AN EVALUATION OF THE VALIDITY OF MULTIDIMENSIONAL SCALING METHODS FOR THE REPRESENTATION OF COGNITIVE PROCESSES

A dissertation
submitted in partial fulfilment
of the requirements for the Degree
of

Doctor of Philosophy in Psychology
in the
University of Canterbury

By

Christopher O. Fraser

University of Canterbury
1980

ACKNOWLEDGEMENTS

I would like to thank the following people for their interest, assistance and support in the preparation of this thesis.

Firstly, I would like to formally acknowledge the valuable assistance of my supervisor, Professor R.A.M. Gregson during my undergraduate and graduate study. His willingness to provide me with constructive advice but with the freedom to pursue my own inclinations were most appreciated.

The following institutions have generously provided me with facilities and material support during preparation of this thesis: the Mathematical Psychology Program, University of Michigan, the Psychometric Laboratory, University of North Carolina, Department of Psychology, University of Newcastle, Centre for Behavioural Studies, University of New England.
I would like to thank James C. Lingoes, Elliot Noma, Clyde H. Coombs, Forrest W. Young, John Hattie, Don Fitzgerald, John Keats and Terry Caeli for valuable discussions on topics relating to this research.
I would particularly like to acknowledge the more personal support provided by Clyde Coombs, Forrest Young and John Hattie.

For the typing of various drafts of this thesis, I would like to thank Vicky Eaddy, Sharyn O'Hara, Kim Blacker and Rhonda Kowalczyk.

I also acknowledge the support of my past and present colleagues at the Gippsland Institute of Advanced Education for encouragement and assistance in providing me with the time to complete this thesis. I particularly thank Warwick Troy and Joseph Kavanagh in this regard.

Finally, I must acknowledge my tremendous debt to my wife Elaina, without whose support, completion of this thesis would have been impossible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Title of Dissertation: An evaluation of the validity of multidimensional scaling methods for the representation of cognitive processes.

Christopher O. Fraser, *Doctor of Philosophy,* *1980.*

Dissertation Supervisor: R.A.M. Gregson,
Professor of Psychology,
University of New England.

This study is an evaluation of the issues involved in providing a meaningful psychological interpretation of multidimensional scaling solutions, for example to regard them as valid representations of the cognitive processes involved in generating the data.

The various metatheoretic approaches that have been developed to define appropriate procedures for the quantification of psychological attributes are discussed and evaluated. It is argued that much current psychological research is based on an inappropriate paradigm. In particular it is argued that the emphasis on magnitude estimation to generate psychological data is misplaced and scales derived from weak-ordered judgements are much to be preferred.

Extending these arguments to the multidimensional case, it is argued that most applications of multivariate methods in psychology have shown insufficient recognition of the theoretical implications of using a particular technique. The application of any method of data analysis such as multidimensional scaling is only appropriate if it can be shown that the assumptions implicit in the scaling model are satisfied for that set of empirical data. In addition some variations in the scaling model, such as subjective metrics scaling, involve additional assumptions which need to be explicitly formulated and tested.

These metatheoretic limitations, as well as evidence on the frequent occurrence of violations of its basic assumptions suggest that multidimensional scaling configurations can at best be attributed with only a limited degree of psychological significance. It is suggested that such value as it does possess is limited to the evaluation of non-dimensional structural hypotheses derived from some prior substantive theory.

An empirical example is presented demonstrating that even when there appears an obvious and intuitive interpretation of the dimensions of a MDS configuration, the solution may be completely inappropriate as a model of the underlying cognitive processes. A second example however describes a more appropriate and successful application of multidimensional scaling methodology. A theoretical interpretation of emotion labelling based on Guttmans (1957) facet theory, was shown to be substantially confirmed in the structure of a MDS configuration.

INTRODUCTION

Since the initial demonstration (Torgerson, 1952) of the feasibility of retrieving the spatial configuration implied by a set of interpoint distances, psychologists have been tempted by the theoretical possibilities presented by multidimensional scaling methods.

The interest in the psychological interpretation of MDS configurations was increased by two apparently significant theoretical advances; the introduction of nonmetric scaling methods (Kruskal, 1964a, b) which appeared to overcome the strict linearity assumptions imposed on the data, and subjective metrics scaling (Carroll & Chang, 1970) which appeared to offer a plausible method of incorporating individual differences within a common perceptual space. Other extensions of the scaling model which appeared to increase its flexibility as a psychological modelling procedure included variations in the nature of the distance function, and in the ways individual configurations could be related to a common configuration.

Most recent theoretical treatments of the MDS model have however cast doubt on the utility of all of these apparent methodological advances and have criticised the lack of any attempt to justify the theoretical assumptions implicit in the application of MDS techniques. The recent emergence of the area of 'measurement theory' (Krantz, Luce, Suppes & Tversky, 1971) as a more rigourous foundation of psychological measurement procedures has emphasised the need for theoretical assumptions implied by scaling techniques, both unidimensional and multidimensional, to be explicitly stated and empirically tested.

One consequence of this greater recognition of the theoretical content of data analysis procedures has been the development of techniques whereby structural hypotheses derived from substantive theory are directly evaluated as a consequence of applying a particular data analytic technique. Techniques for confirmatory analysis have been derived for both factor analysis (Jöreskog, 1969) and multidimensional scaling (Bentler & Weeks, 1978, Borg & Lingoes 1978, Bloxam, 1978).

The aim of this study is to evaluate the contribution these various methodological developments in multidimensional scaling can make to the goal of providing an interpretation of a spatial structure derived from data that has some substantive theoretical significance.

Multidimensional scaling will be regarded as a psychological measurement procedure designed to derive simultaneously scale values on more than one dimension. This study will thus attempt to apply the perspective of the measurement theoretic approach to unidimensional scaling to the more general multidimensional case. Considerable attention will thus be paid to the metatheoretical aspects involved in the quantification of psychological dimensions, and Part I of this thesis is devoted to this topic. The first chapter, which deals with various conceptualizations of the logic of unidimensional measurement, is an extended version of a paper recently published in the <u>British Journal of Psychology</u> (Fraser, 1980).

Part II is a review of all significant methodological variations in multidimensional scaling procedures and an evaluation of the extent to which meaningful psychological interpretations can be attributed to different types of scaling solutions.

Part III describes some empirical research designed to demonstrate what appears to be the most promising approach to the utilisation of MDS methodology in the development of psychological theory. This section also demonstrates how incorrect conclusions can easily be drawn by an uncritical approach to the interpretation of MDS solutions. Chapter V in this section is a more extensive treatment of a previously published study (Fraser, 1976) and a copy of this paper is also attached.

Chapter VII summarises the approach to the utilisation of MDS in psychological research that has been developed within this thesis, and compares it with the position of other major theorists on this topic.

PART ONE

THEORETICAL ISSUES IN QUANTIFICATION IN PSYCHOLOGY

CHAPTER ONE

THE LOGIC OF MEASUREMENT AND PSYCHOLOGICAL THEORY

The increasing use of quantitative methods in psychology should imply an increasing emphasis on both methodological and theoretical aspects of obtaining numerical measures of attributes. While psychology has developed an increasing number of methods to quantify attitudes, abilities, needs, tendencies etc., very little attention has been devoted to the logical justification of such procedures. Psychology has by and large relied on a set of essentially ad hoc statistical techniques, subsumed under the titles 'reliability' and 'validity', to justify its measurement procedures. That is, it depends on establishing a system of interrelationships among a variety of measures, rather than an independent justification of each one singly.

Most texts on psychological measurement have worked within this paradigm to develop practical procedures for scale construction and evaluation (e.g. Nunnally, 1967, 1970; Cronbach et.al., 1972; Lemke and Wiersma 1976). As Nunnally states "the ultimate test is how well the scales that are derived fit in a nexus of lawful relations with other scales" (1967, p.34).

A more firm foundation for psychological measurement however, is clearly provided by a consideration of measurement from a fundamental point of view, that is, independent of other scales. The idea of fundamental measurement has until recently, only existed within the confines of classical physics. The rather restrictive set of properties that have been derived by physical theorists to represent the conditions under which fundamental measurement can occur clearly preclude most psychological attributes. It is hardly surprising then that psychological measurement has long been regarded as quantitatively and qualitatively of a lower order than physical measurement.

There appears to have been two main consequences of this for the theory and practice of measurement in psychology. The first has been the use of a less deterministic approach to measurement, regarding scales as approximate indices only, subject to an unknown amount of measurement error. The second has been the concept of lower order scales embodied in S.S. Stevens' analysis of measurement theory (Stevens, 1939, 1946, 1951, 1968).

Recent developments in psychological measurement theory have however, demonstrated that the concept of fundamental measurement can be extended in a way that is appropriate for psychological attributes and have identified some substantial logical flaws in previous approaches. By reviewing these developments it is hoped to show that much of the currently accepted theory and practice of psychological measurement is in need of a substantial revision.

Classical Physical Measurement Theory

In physics, as in psychology, an explicit formulation of the logical conditions for measurement came long after the development of practical measurement procedures. The first attempt to provide a logical justification for measurement was given in the works of Helmholtz (1887). Helmholtz introduced the notion that fundamental measurement depended on the presence of an empirical operation to compare directly the quantities of objects (e.g. juxtaposition of length, weights in a pan balance), and attempted to identify the conditions under which such an operation could lead to the development of a measurement scale. This was developed at a more formal level in Hölder's (1901) classic theorem specifying the axioms of quantity. These specified the relations that must hold between entities for them to be used as elements in additive and multiplicative functions, i.e. as quantities, and also that one be able to add together two quantities to make a new quantity.

The consequences of this for the process of measurement were given their most complete exposition in the works of N.R. Campbell (1920, 1921, 1928, 1938). Campbell saw measurement as the demonstration of an isomorphism between the ideas of quantity and the magnitudes of the property to be measured. (A magnitude is the amount of a specific property possessed by a specific object). The way to do this was to demonstrate that the magnitudes obeyed the axiom of quantity developed by Holder.

Measurement thus depended on being able to observe relations between physical objects as a consequence of performing an empirical operation. This empirical operation was the defining operation of the measurement. There was no meaning to the concept of two objects being equal with respect to some quality without first defining the operation used to test equality.

Because the axioms of quantity required an operation of addition as well as equality an empirically demonstrable capacity for addition was regarded as a fundamental importance to the measur ability of attributes. Campbell and several subsequent theorists (e.g., Cohen and Nagel, 1934, Ellis, 1960) regarded the existence of an empirically defined concatenation operation as practically essential for fundamental measurement. Campbell referred to this as the additive property, although it is now usually denoted as the extensive property.

Campbell did allow for a potential theoretical distinction between extensive and fundamental measurement, but did not see it of any practical importance since alternative methods were unlikely to be available. Ellis (1966) in fact uses the terms synonymously, using the term direct to indicate fundamental in the sense we have defined it here, i.e., independence from other measures.

Clearly not all physical scales possess this presumed fundamental extensive property. Density for example, possesses an operation for ordering objects (which floats on which) but not for adding them. Density and temperature are in fact examples of what are known as intensive properties, since combining objects produces the average, not the sum of their densities. Clearly several important physical properties were intensive but were still capable of being measured successfully by virtue of known mathematical laws relating them to other properties; for example, density as the ratio of mass to volume. To incorporate this Campbell introduced a secondary type of measurement, termed derived measurement, for attributes which could be expressed in terms of a mathematical relation with extensively measured attributes.

Evaluation of Classical Physical Measurement Theory

The classical physical view of measurement theory incorporates two critical features, both of which have been criticised by subsequent theorists as being overly restrictive.

The first is the very restricted view of what can constitute measurement. Only one type of scale was countenanced; 'quantity' had one set ofproperties and these were strictly not negotiable. Magnitudes had to meet these in their entirety, or they could not be measured.

The second is the view that the relations implied by the concept of quantity must be directly exhibited by actual physical objects; that is, no distinction was drawn between the object, and the measure assigned to that object. Since these relations included the concept of addition this implied that it must be meaningful to talk of the addition of objects.

This for example was the basis of the conclusion of a committee of which Campbell was a member, reporting on whether sensations in general, and the sone scale of loudness in particular, constituted measurement. In its Final Report (Ferguson et.al., 1940) it concluded that:

> 'any law purporting to express a quantitative relation between sensation intensity and stimulus intensity is not merely false but is in fact misleading unless and until a meaning can be given to the concept of addition as applied to sensation'.

This insistence that relations can only be exhibited between physical objects clearly poses problems for the concept of derived measurement. No external meaning can be given to statements such as 'twice as dense','twice as hot', at least in this direct sense of adding two objects with the particular density, temperature, etc. This fact was criticised by B. Russell in Principles of Mathematics. As Russell states, one can only add quantities not magnitudes, "addition of two magnitudes yields two magnitudes, while the addition of two quantities does give a new single whole" (1937, p.178)

Classical Psychological Measurement Theory

The measurement theory of S.S. Stevens has gained widespread acceptance within and outside psychology, and as the first comprehensive attempt to develop a theory applicable to psychology might reasonably be regarded as the classical psychological approach.

Stevens' main innovation was to remove the restriction that numbers assigned as measures had necessarily to obey the laws of quantity, and introduced the possibility of other types of scales. Instead measurement was defined broadly as "the assignment of numbers to objects or events according to rules". (1951, p.1), with the problem being transferred to identifying the type, rather than the existence of measurement. This classification of the type or level of measurement was to be done by identification of the type of mathematical transformations that left the scale form invariant (i.e., so it still obeyed the rule).

Stevens system initially recognized the existence of four different levels of measurement; nominal, ordinal, interval and ratio. The level of measurement of a scale implied in general what sorts of statements or conclusions could be derived from it, and in particular what type of statistical test should be used with it.

Evaluation of Classical Psychological Measurement Theory

Stevens-tradition (ST) measurement theory clearly provided some advantages over the classical physical approach. It drew attention to the distinction between the formal system (numbers) and the empirical system it represented and the need for the properties of each to be logically connected. It also identified the arbitrariness of insisting on using the properties of the number system in their entirety as the only possible formal system that could constitute measurement.

The critical deficiency in Stevens' system is that while apparently recognizing the distinction between the formal and empirical systems, it does not in fact characterize them separately and identify the nature of the connection between them. Ross (1964) argued that because Stevens did not explicitly state how the assignment of scale values related to specific empirical operations, that he in fact implicitly incorporated non-falsifiable assumptions into this process. Ross drew attention to an inconsistency in Stevens' approach between his theoretical statements and the practical procedures he advocated. For example concerning the formal conditions for measurement he states:

> "Measurement is only possible because there is a kind of isomorphism between (1) the empirical relations among properties of objects and events and (2) the properties of the formal game in which the numerals are the pawns and operators the moves" (Stevens, 1959, pp.20-21)

However, many of the practical applications Stevens has quoted imply that he treats this isomorphism as axiomatic, something which must be present if measurement has been achieved, rather than a condition that needs to be empirically established. For example, the technique of magnitude estimation, which assumes subjects can directly generate numbers in proportion to their apparent magnitude, contains no implications of even 'a kind of' isomorphism between basic empirical operations and the number system.

This inconsistency between theory and practice can also be identified in the procedures used for the determination of scale type. Stevens maintains that the classification of scale type can be accomplished by the principle of scale invariance, without reference to any relationship to empirical events or other scales. He states:

> "We may seek the final and definitive answer in the mathematical group structure of the scale form; in what ways can we transform its values and still have all the functions previously fulfilled". (Stevens, 1951, p.29)

However, as has often been pointed out (e.g., Ellis, 1966; Nunnally, 1967; Prytulak, 1975; Fraser, 1980) this definition does not correspond with his actual practice in classifying scale types. For example, many psychological measures are obtained as frequency counts; number of trials, bar presses, words recalled and so on. And counting, as Stevens agrees is a ratio scale. Yet Stevens still insists that "most of the scales used widely by psychologists are ordinal scales" (1951, p.20). I.Q. for example, is an ordinal scale according to Stevens, yet by his definition, it could be converted to a ratio scale by counting questions. Clearly something else is being taken into account, although it is not specified what it is. Ellis (1966, p.63) concludes that Stevens "simply makes the classification as though the reasons were self-evident", and Nunnally (1967, p.21) that "no one has made it clear what types of evidence would justify the assumption of a particular scale type".

Ellis (1966) demonstrated that an inadmissable transformation of a scale could still serve all of the functions and retain all the information of the old scale. For example, the application of a log transformation to the Kelvin scale for temperature is clearly inadmissible, yet could still be used to formulate physical laws to relate temperature to other scales. The laws would not be of the familiar form, but then these only exist because of the particular type of scale used to express them in.

Clearly then the principle of transformational invariance as expressed by Stevens is logically insufficient as a means ofdeterminingscale type. Prytulak (1975) has suggested that it can be made sufficient by basing it on empirically determined relationships with other scales. To support this he observes that the procedures used by ST theorists do in fact implicitly, if not explicitly, consider transformations to other scales. They avoid the direct recognition of the second scale by using examples such as subjective length where some familiar scale is implicitly present (i.e., physical length) and so the reader can readily appreciate the relation between the scales without any overt recognition that a transformation is involved.

So in fact, as Prytulak observes, the actual choice of scale type results from a consideration of all such implicit combinations that are relevant to the use of the scale, and involves such intuitive criteria as the relative frequency, importance and range of situations in which transformations of the higher and more desirable types occur.

While there is undoubtedly a role for this extended version of ST measurement theory in dealing with the psychological measurement procedures which draw their justification from established relationship with other scales, it is a step back from the objective of deriving fundamental measurement procedures in psychology. The transformational properties of various scales are of some use as a purely descriptive classification, but they cannot be regarded as constituting a philosophy of measurement.

Axiomatic Measurement Theory

An alternative line of development in measurement theory was to return to the concept of fundamental measurement as developed in the classical physical approach, but to extend it to include the situations more commonly found in psychology. This approach recognizes that there is no logical justification for the restricted notion of fundamental measurement proposed by Campbell, which regarded it as equivalent to extensive measurement.

A more realistic definition of fundamental measurement should incorporate the extension proposed by Stevens of including less than the full range of properties of the number system, while still retaining the need for an empirically established correspondence between the empirical system and these properties. Coombs, Raiffa and Thrall (1954) urged such an extension to include scales based on any type of formal system, including those such as lattices and partial orderings which cannot be represented numerically in any natural way.

The earliest approaches attempted to characterise the precise properties in the empirical system required to justify ordered scales. Hempel (1952) for example, derived the axiomatic conditions for ordinal scales based on what he termed 'quasi-series.' These required the fairly strict conditions that both an equivalence and an ordering relation could be empirically observed in order to generate the series. Luce (1956) proposed as a natural and more realistic extension the concept of 'semiorders'. These allowed for the more realistic situation in psychological ordering where the indifference relation is not transitive. That is, while one may be indifferent between A and B, and between B and C, one may not necessarily be indifferent between A and C, as the first two differences may be less than one jnd, while the combined differences exceed it.

A more general formal analysis of the process of deriving the precise conditions required to justify a measurement scale of a particular type was given in Scott and Suppes (1958) and developed more extensively in Suppes and Zinnes (1963). They identified the two fundamental problems of measurement as (1) the justification of the assignment of numbers to objects or phenomena, and (2) the specification of the degree to which this assignment is unique.

The first fundamental problem was stated in more formal terms as the need to:

> "Characterise the formal properties of the empirical operations and relations used in the procedure and show they are isomorphic to appropriately chosen numerical operations and relations"
>
> (Suppes and Zinnes, 1963, p.4)

The precise nature of the isomorphism, that is, what relations defined in the formal system are assumed to represent relations holding in the empirical system, constitutes the measurement model. The proof that a numerical (or other) assignment can be made to a given empirical system to produce a representation of the type defined in the measurement model is provided by the representation theorem. This is based on a specification of a number of axioms which represent the minimum conditions that must be satisfied by the empirical system in order for a scale of that particular type to be produced.

The uniqueness theorem specifies the type of representation that is obtained by determining the type of transformations that are permissible for a given numerical assignment, while still obeying the axioms of the representation theorem. It is thus analagous to Stevens' admissible transformations principle, except the condition of scale invariance has been more explicitly defined.

Suppes and Zinnes derived representation theorems for a variety of previously defined measurement models, including quasi-series, semi-orders and extensive systems. They also developed theorems for a class of measurement models that produce difference scales. These were based on a quaternary ordinal relation in the empirical system corresponding to relations on differences between pairs of scale values.

Further classes of measurement models have been proposed by subsequent theorists and include for example bisection systems (Pfanzagl, 1968), conjoint measurement (Luce and Tukey, 1964), probability representations (Krantz et.al., 1971), Grassman structures (Krantz, 1974). A comprehensive review of most varieties of psychological measurement models has been provided in Krantz et.al., (1971). The details of these will not be considered here. Instead the concept of derived measurement will be reconsidered in the light of the extended definition of fundamental measurement.

Derived Measurement

Pfanzagl (1968, p.31) has maintained that derived measurement is not really measurement at all, being simply a combination of other scales. He argues that only fundamentally measurable qualities should be used in the search for empirical laws.

This is an overly restricted view, however, as it ignores the importance of the total set of interrelationships that exist among the scales in a given system. In systems where the interdependence of scales is high the number of scales which must be independently measured is correspondingly reduced. The physical system is in the happy situation of having a high interdependence of scales, together with a large proportion of extensively measurable scales. In fact, only six independent dimensions are needed to describe completely all aspects of the system, although it is convenient to use far more. All other scales however can be expressed as simple functions of a set of six basic dimensions. Since all physical scales are ratio scales, that is completely determined except for an arbitrary choice of unit; we need only specify the units for six basic dimensions to determine the whole system. Several writers have emphasised the theoretical importance of the scales chosen as the basis of a dimensional system. Campbell (1920) called these basic scales, Palacios (1956) defined them as fundamental and Ellis (1966) called them independent scales. We will follow the terminology of Krantz et.al., (1971) and denote the scales in the base as primary, those that are not, as secondary. This distinction is somewhat arbitrary as the choice of base is not unique, but does indicate the importance of deriving scales from others within a system, even when an independent fundamental measure may be available. (Campbell used the term quasi-derived for scales in this situation; 1920, p.379). Hempel (1952) also referred to the importance of systematic relationships in distinguishing stipulative and vicarious as two categories of direct measurement. Stipulative is an arbitrarily defined relation with another measure, whereas vicarious measures arise as a consequence of empirical laws. Krantz et.al., (1971) however give a more accurate account of the different types of derived measurement. Attributes which are defined as a simple function of two other attributes, such as density as the ratio of mass to volume, might appear to merely represent a more convenient expression of this function. They, in fact however, represent laws of similitude, i.e., statements of the similarity of certain classes of physical systems. The derived measure is thus not arbitrary, but a measure of the constant value defined by the law. As Krantz et. al., pointed out, the basis of laws of similitude is that two extensive attributes, e.g., mass and volume, both depend on the same empirical concatenation operation. The non-extensive attribute, density, arises as a product of the law.

A second type of derived measure results from laws of exchange. These also relate two extensive attributes to a third attribute, but in this case the two concatenation operations are independent. They are related by a physical law of conservation in a closed system. For example P/VT= constant, and thus a measure on two defines the third. These laws thus depend on an empirical compensation of concatenation operations, rather than an equivalence.

Krantz et. al., (1971 , p.488) discuss the relation between derived measurements of this kind and the concept of conjoint measurement. This is a type of fundamental measurement based on objects considered as a conjoint effect of two or more components. Axioms have been developed specifying the conditions under which this can lead to the construction of a numerical representation that is additive over those components which does not depend on the existence of a concatenation operation for the objects.

Laws such as m=Vd can thus be regarded as establishing a conjoint scale on one attribute e.g., density, conditional upon the law relating it to the other two being true. Since this measurement can be defined as fundamental the argument that derived non-extensive scales are not measurement clearly cannot be maintained.

In some cases derived scales may also be capable of being extensively measured. This is, we may derive two scales, an extensive and a conjoint one. In this case they prove to be power functions of each other. Krantz et.al., (1971) derive several other relations which must hold when two or more types of measurement hold simultaneously (1971, ch.10).

These interrelationships between different types of measures and the somewhat arbitrariness of the choice of measure to form the basis to the system, show that the distinction between fundamental and derived measurement is more a matter of convenience than of fundamental theoretical importance.

Evaluation of Axiomatic Measurement Theory

It must first be recognized that axiomatic measurement theory only specifies the conditions under which certain types of scales are theoretically possible. It does not provide a procedure for obtaining such a scale, or even show that such a procedure is feasible.

The evaluation of scaling procedures under this approach will rarely be completely definitive as the conditions required by the measurement model will hardly ever be completely satisfied. There will thus be some doubt about whether these are systematic departures casting some doubt on the measurement model or simply errors due to noisy data.

Criticism of the impractibility of measurement theory are given by Luce (1972) and Falmagne (1976). Luce argues that it is unlikely that psychology could develop a system of psychophysical measures analogous to fundamental measurement in physics due to the lack of any firmly established invariant relationships between them. Falmagne (1976) comments that measurement theory has not yet demonstrated any constructive impact on behavioural research. He also suggests that the algebraic format of the existing theories does not represent an appropriate way to accommodate error. Such a deterministic model does not provide an appropriate statistical basis for testing for violations. He proposes instead the development of measurement theories based on random variables, rather than constant values and outlines an example of a 'random conjoint measurement theory'.

However, despite these limitations it must be a distinct advantage to provide for every psychological scaling procedure an explicit statement of the conditions on which measurability is based. As Pfanzagl has observed "there is a great number of procedures for computing scales for which no definition is available, except the one implicit in the computational procedure" (1968, p.9).

Probably the major advantage of the axiomatic approach to measurement is that it explicitly recognizes the role of theory in measurement. Measurement is seen as an integral part of the theory, rather than simply a preliminary stage prior to theory construction. As Coombs (1964) has remarked:

> "all knowledge is the result of theory - we buy information with assumptions - "facts" are inferences, and so also are data and measurements and scales" (1964, p.5)

A consequence of this has been a movement towards theory construction where measurement questions are seen as an integral part. The advantages of this approach have been emphasised by Anderson (1970) and Krantz (1974). As Krantz states:

> "at the present stage of development of quantitative theory in social science it is impossible to separate the search for interesting empirical laws from the discovery and refinement of measurement procedures" (1967, p.13).

Summary

The arguments presented here should provide conclusive evidence that many of the traditional attitudes to psychological measurement which have assigned a low priority to its theoretical content are no longer justified. In particular, the dominant theoretical conceptualizations that have been held to structure these ideas are inadequately formulated and inconsistently applied. As Cliff (1973) observed in a recent review of scaling procedures:

> "The recent achievements in measurement theory provide hardly less than the basis for a revolution in the definition of psychological variables" (1973, p.447).

CHAPTER TWO

MULTIVARIATE ANALYSIS AND PSYCHOLOGICAL THEORY

Most descriptions of the various methods of multivariate analysis take as their starting point an array of numbers, usually regarded as being structured as cases by variables, and describe various procedures for transforming this 'data matrix'. The solution thus obtained is seen as a method of summarising or representing the original matrix, and the procedure is evaluated in terms of the statistical relations between the two.

This approach thus regards multivariate analysis as a branch of statistics, to be applied to a particular content area, in this case psychology. To then attempt, as is often done, to accord theoretical significance to the representations arising from these multivariate procedures is clearly inconsistent with the logic of psychological measurement. That is, the critical question whether multivariate analyses represent psychological models or simply methods of data reduction can scarcely admit of the former response as long as statistical rather than psychological criteria are used to derive them.

As detailed in the previous chapter, measurement is the process of deriving a representation of certain properties of the data, and the theoretical assumptions made about them. A two stage process of restricting measurement questions to the stage prior to the actual analysis, and even then placing little emphasis on it, can thus scarcely encourage an appropriate interpretation of the role of multivariate methods in theory construction. It seems not unreasonable to attribute the inappropriate attitudes regarding multivariate procedures implicit in many of their past and current applications to this historical separation.

For example, factor analysis was at one stage a very popular and extensively used technique in psychological research and much confidence was placed in the theoretical significance of the factors that resulted. However, the emergence of a number of technical problems relating to the method showed that this confidence was misplaced, and that a more cautious approach to interpretation was required. Current attitudes to its use now range from outright rejection to a continued insistence on its theoretical significance.

In between are attitudes such as those expressed by Royce (1973) who favours its use in what philosophers of science call 'low level theorizing' i.e., conceptualizations based on empirical regularities but not regarded as complete and definitive. For example, he states (1973, p.3) that (factor analysis) "can provide us with some potentially powerful theoretical constructs, and possibly some clues regarding how these constructs are taxonomically arranged, but the building of theoretical structures which elaborate on relationships between variables and otherwise 'explain' observables is an extra-factorial enterprise".

However, all these attitudes miss the essential point which is, as expressed by Krantz (1974, p.171) that 'psychometric techniques are not theoretically neutral; rather they impose a theoretical structure whose psychological content deserves analysis', A more logical approach then is to view measurement as an integral part of the application of psychological theory to data, and therefore something which relates to the whole analysis, rather than something which must be 'struggled with prior to theory construction' (Krantz, 1971, p.171). This view is embodied in the conceptualization of multivariate analysis as the application of a measurement model. As with unidimensional measurement, multivariate analysis derives a representation of the data by means of a set of theoretical assumptions relating properties of the data to aspects of the representation. This representation will usually, but not always, be multidimensional, i.e. a set of numerical scales. Spatial models are thus a direct extension of unidimensional measurement, and non-spatial models, such as clustering, can logically be considered as further generalizations of the measurement process.

This approach of viewing multivariate analysis as measurement which arises as a consequence of theory provides a more appropriate basis for the validation of such procedures. An integral part of every theory is that it specifies which empirical results should be observed in order to determine whether the theory is supported. Therefore, viewing multivariate analysis solutions as representations implies that they make specific predictions about the empirical system they represent. Their justification should consist therefore of identifying and evaluating these predictions.

Scaling And Data Theory

It is clear then that much more consideration must be given to the types of observations that should be used as input to multivariate procedures. That is the data analysis consists of a model or formal system, and the observations an empirical system. The measurement model contains the set of theoretical assumptions specifying a correspondence between the formal and empirical systems, that is what relations defined in the formal systemare assumed to represent relations holding in the empirical system. The measurement model however only specifies the conditions under which a scale or scales of a specific type can be derived. The actual process of scaling, i.e. assigning numbers to objects or properties is a separate enterprise.

The terms scaling and data analysis as we have used them here are essentially equivalent. That is ideally data analysis can be defined as deriving scales under a set of assumptions specified in a measurement model. Certainly in the unidimensional case no useful distinction can be made. However where more than one scale is to be derived this ideal requirement is not always observed. That is, many methods of data analysis operate on sets of unidimensionally derived scales, for example the correlational based procedures such as factor analysis, discriminant analysis, etc. The term multidimensional scaling is thus logically defined as any procedure based on a measurement model which provides for the derivation of more than one scale simultaneously. It is this class of procedures of data analysis that is the prime focus of this thesis, although much of the discussion is also relevant to other multivariate methods.

In all cases one needs to consider the means by which a correspondence may be established between a set of observations and a numerical scale. On what basis can we attribute the relational properties required to produce numerical scales for different types of behavioural observations? Most conceptions of this have been developed within the framework of psychophysical theory. While this term is strictly appropriate only in situations where there is a known physical scale, the principles can logically be extended to cases where it is not.

Modern psychophysics make a formal distinction between four qualitatively distinct judgements; detection, discrimination, recognition and magnitude estimation. (Luce, Bush and Galanter, 1963). These judgements clearly do not have an equivalent psychological status. Detection and discrimination judgements can be evaluated against a known physical status, and thus false positive statements can be detected. However for magnitude assessment techniques there is no independent physical specification of what the appropriate response should be.

One can identify a number of quite distinct theoretical positions over the measurement properties of these psychophysical methods. These have led to a number of quite distinct bodies of theory and data regarding the procedures for the construction of psychological scales. One of the most important distinctions relates to assumptions about the way internal psychological effects (usually expressed as 'sensation magnitudes') are quantitatively related to external observable responses. Two contrasting positions can be identified here. The first states that the sensation evoked by a stimulus is essentially a phenomenological event which can be directly experienced by the subject and perfectly represented in his judgements.

The alternative position is to regard sensation as a theoretical construct which relates to the internal representation of stimuli. The observed responses of the subject are assumed to be based on these internal representations and derived from them by means of some form of cognitive process. That is, this represents the cognitive modelling approach where we do not assume that the empirical system we seek to explain is directly accessible and perfectly represented by our observations. When a subject is exposed to a set of conditions we observe not the direct effect of these conditions but their effect mediated by some internal psychological processes and the processes by which this is transformed into an observed response. The aim of this approach therefore is to develop a plausible model of these processes which then enable the form of the internal representations to be derived which best account for the data.

Of course a number of possible models can be proposed and the meaning of the term sensation will thus vary according to which is being assumed. This leads us to a third compromise position which rests on more pragmatic parsimonious grounds. The term sensation is assigned to a scale only if it is generalisable across a wide range of situations. This point was expressed by Luce, Bush and Galanter (1963, p.207):

> "If a scale serves no purpose other than as a compact summary of the data from which it was calculated, if it fails to predict different data or to relate apparently unrelated results, that is, if it is not a theoretical device, then it is worth but little attention and surely we should not let it appropriate such a prized word as 'sensation'. If however a scale is ever shown to have a rich theoretical and predictive role then the scientific community can afford to risk the loss of a good word".

Consistency across methods must surely be a prime consideration for meaningful scales. However before subscribing to this cautious approach we should consider the arguments for the positions taken by the first two approaches regarding the essential nature of the sensory continuum.

The first position is most closely identified with the work of S.S. Stevens and his colleagues. Stevens' position was that one could treat numerical responses to stimuli as direct numerical measures of attributes. Ross (1964) however drew attention to some ambiguity in Stevens writings over whether this meant acceptance of a phenomenological or behavioural approach to psychology. He compares passages quoted from Stevens (1956) which appear to indicate contradictory assumptions regarding the relationship between the subjective numerical estimates and the concept of sensation magnitude. On one hand he states:

> "The purpose of this study was to try to develop and refine a method for the direct quantitative assessment of subjective magnitude".

and also that:

> "The problem in subjective measurement is to arrange the conditions and the task in such a way that O can assess his impressions and communicate them to E in a quantitative language with as few biasing cues, suggestions and constraints as possible".
>
> (1956, p.2)

These quotes appear to imply that 'sensations' are the basic elements of the object language of psychology, and the psychologist and the subject have to work in tandem to provide a way to measure these sensations. The subject's statements are thus no longer private statements about sensations, but are on a par with the statements of another psychologist, and are thus part of the metalanguage. That is, the subject is acting as a scientific observer of his own sensations, even though no one else can observe them.

However, this point of view appears contradicted by the statement in the same paper:

> "By a scale of subjective magnitude we mean quantitative scale by which we can predict what people will say when they try to give a quantitative description of their impressions".
>
> (1956, p.2)

This statement retreats to a purely operational definition of subjective magnitude, equating it with the subject's numerical responses.

It is beyond question that subjective responses should be part of the object language. This view has been strongly urged by Bergmann and Spence (1944) with explicit reference to this type of psychological scaling procedure. However the second interpretation above implies that the term 'sensation' is to be applied to the sensory scale underlying one specific class of judgements. As a justification for this it has often been claimed that magnitude estimation and ratio scaling are 'direct' scaling procedures while those based on discrimination are 'indirect' methods. Stevens (1957) argued that the reason scales based on category judgements were not linearly related to magnitude estimation scales was that subjects in a category rating task confused discriminability with psychological distance. The arbitrariness of this distinction has been criticised by several authors (e.g. Treisman, 1964; Zinnes 1969; Krantz, 1972 b; J.C. Falmagne, 1974) and Stevens' later writings appear to accept the position that there is nothing particularly unique or direct about the magnitude estimation task, and he refers to it simply as a matching task, involving the matching of numbers to some other physical variable (Stevens, 1964, 1966 a,b,c).

On methodological grounds the evidence for using magnitude estimates as stable psychological measures is somewhat equivocal. An enormous number of studies have attempted to demonstrate the consistency of these estimates, usually by demonstrating that they lead to the well known power law form of the psychophysical function. Zinnes (1969) in a review of these studies concluded that the results were generally unconvincing. Although a power function was often found to give a reasonable fit the value of the exponent clearly varied with experimental conditions. Zinnes also pointed out that the level of statistical analysis, usually a least-squares fit of straight lines to log-log data, has limited power to discriminate between different functions. Other reviewers (e.g. Eisler, 1965; Luce, 1972; Krantz, 1972 a) have concluded that the regularities of the data are sufficient for it not to be rejected on these grounds. Krantz (1972 a) cites four generalizations of data arising out of Stevens methods (in addition to the power law) which he regards as consistent enough to require explanation. These relate to consistencies between and within the three classes of procedures which Stevens refers to as 'ratio scaling'. (Krantz objects to this term and insists on the distinction between this and the concept of a ratio scale). The three methods are magnitude estimation, where the stimuli are presented one at a time and the subject 'assigns numbers in proportion to the sensations evoked' on the attribute used to define the sensory continuum, ratio estimation where two stimuli are presented and the subject is asked to judge the sensation ratio, and cross-modality matching where stimuli are presented one at a time and the subject chooses a stimulus in another continuum that matches the given stimulus.

The four consistencies Krantz identifies are:

(i) that magnitude estimate data with different moduli differ by a constant multiplier (a modulus is a constant stimulus assigned a predetermined value).

(ii) That ratio estimates behave like numerical ratios. That is if we denote the estimated ratio of a to b as $M(a,b)$ then we should have $M(a,b)/M(b,c) = M(a,c)$

(iii) The ratio estimate is equal to the ratio of the magnitude estimates for members of the pair

(iv) Cross-modality estimation is consistent with magnitude estimation on the separate continua. That is, if stimulus x_i from dimension X is matched with stimulus y_i from Y, where the constant match for the cross-modality matching is (x_k, y_i), then the magnitude estimation of y_i and the magnitude estimation of x_i should be the same ratio of the magnitude estimation of their respective moduli, $(y_i, x_k.)$

The last of these has been the most extensively researched, mostly by attempting to demonstrate consistency of the power function exponent derived from the two methods (with the aim of thus establishing the generality of the power law). Stevens (1964) concluded that the exponents generally showed a range sufficiently close to indicate that the evidence was converging. Mashour and Hosman (1958) however disagreed claiming that a large percentage of results are sufficiently different to regard the power law as unsatifactory even as a first approximation.

However, it is on theoretical rather than empirical grounds that the most serious objections to direct magnitude estimation are made. Both Luce (1972) and Falmagne (1974) criticise the fact that data based on discrimination and detection cannot be reconciled with this approach.

A more philosophical objection is that the approach conceptualises the sensation dimension as being intrinsically numerical. Krantz (1972) argues that it is highly implausible to regard sensation as possessing the inherently numerical properties that would be required to account for the ratio-like consistency of cross-modality matching. He argues that it is far more reasonable to regard numerical estimation as a second stage process following the initial judgement.

Wagenaar (1975) supports this criticism and demonstrates that if a transformation to the number system from the subjective magnitude scale is allowed then it becomes impractical to distinguish the Stevens and Fechner forms of the psychophysical function. While the identification of this is no longer regarded as a critical question it does cast doubt on the use of evidence about the power function to indicate stability of psychological scales.

A central question to the present investigation is the relation of the scales produced by various methods to the concept of measurement. Magnitude estimates could be justified as fundamental ordinal measurement if the assumption of the Shepard-Krantz theory were accepted. Similarly the parameters of the probabalistic models can be considered as derived measures (from the extensive measure probability). However, as Luce (1972) observes it seems unlikely that either of these will prove sufficiently general to serve as the basis of a system of psychophysical measurement.

One seems to be presented with two options. One is the approach recommended by Luce (1972) of regarding the human observer as a measurement device (i.e. a system for transforming one attribute into another in an approximately one-to-one manner). The task of the psychophysicist is to identify the processes operating in the measuring device.

The alternative is to attempt to derive scales which are fundamentally measurable but to restrict the empirical relations required to those involving minimal assumptions about the nature of internal processes, e.g. ordinal judgements. Zinnes (1969) in recommending that the theory of scaling should become the theory of choice urges the need to avoid the untenable assumption that subjects are measuring instruments capable of being activated at any level of measurement by appropriate instructions.

The progress in measurement theory has demonstrated that a very firm basis for the foundations of psychophysics can be provided by weak ordered cognitive judgements i.e. responses to questions like 'Is stimulus a at least as intense as stimulus b?' Feasible scaling procedures from judgements of this type are provided by models such as Coombs unfolding analysis (1950) and Guttman scaling. Direct magnitude estimates do offer some advantages of convenience and one cannot reject such data as entirely lacking in validity. It is clear however that direct numerical measures should be treated with some caution in situations where some theoretical relevance is to be attributed to the results. There is certainly no justification for the position that numerical estimates represent direct measures while categorical data are indirect. One can certainly endorse Zinnes position over the need to abolish this distinction. As Zinnes states (1969, p.463)

> "There is no important sense in which numerical responses are more direct than non-numerical responses. If anything numerical responses are more indirect because they are more complex, depending not only on perceptual processes, but on the subject's use and familiarity with numbers as well. Ultimately a theory of numerical responses will have to be more complex than a theory of non-numerical responses because it will have more to explain."

Data Analyses as Models of Cognitive Process

The extension of the ideas of measurement to the multivariate case implies viewing multivariate procedures as formal models of the behaviour used to generate the data. What this means in terms of the psychological meaning that might be attributed to the results depends on the sense in which we use the term 'modeling'. The term model can be used in a limited sense to refer to any mathematical system used to predict some quantified aspects of behaviour. The aim in this case is to provide a parsimonious description of the structural relationships in the data, without any claims to identifying psychological reality. Such an approach to modelling can be demonstrated in the mathematical models of learning developed in the 1950's by W.K. Estes, and others (Bush and Mosteller, 1955; Estes, 1959; Luce, 1959).

The emphasis in these initial approaches was on deriving accurate predictions of certain aspects of the learning process under various conditions; they were seen as formal systems accounting for the data rather than characterizations of 'true' mechanisms. The utility of such a perspective depends on philosophical orientations to science and psychology which it is not proposed to consider here. Rather it will simply be stated that this positivist approach of equating scientific explanation with scientific description will not be regarded as useful here. It will be assumed that the aim of psychology is to describe the cognitive processes that gives rise to behaviour itself. So the term model will be used in the sense of a quantitative theory, an attempt to specify precise assumptions about underlying processes.

An important distinction can be made between two classes of formal models that relates to the centrality of assumption about causal processes. Deterministic models specify the response uniquely for given antecedent conditions, while stochastic models specify response strategies in probabalistic terms. While not in itself an essential difference (the deterministic model is just a limiting case of the stochastic model) it implies different strategies in terms of the search for psychological relevance. Deterministic models stress the fine-grain prediction of behaviour in a specific situation. The models thus tend to be expressed at the level of detail and comprehensiveness, and in the same terminology as our psychological understanding of the phenomena in question. That is, rather than searching for abstract simplifications behaviour is portrayed in its original complexity. The implication that they are direct parallels of actual cognitive processes is therefore obvious. Examples of this approach are provided in the rapidly developing area of computer simulation. Models of a variety of cognitive processes have been developed, for example the information-processing models of verbal learning (Feigenbaum, 1961), and of problem solving (Newell and Simon, 1972; Simon and Newell, 1974).

Stochastic models on the other hand tend to predict behaviour at a more gross level of approximation. The emphasis is more on demonstrating the generality of simple formal models across a wide range of experimental situations. The earliest examples of the application of mathematical models to psychology, the learning models referred to before, used stochastic models with probability of a response as the main dependent variable. While these were initially considered in a completely behavioural sense as formal models, derived by a curve-fitting operation, which were capable of explicitly derivable predictions about behaviour, more recent theorists have viewed them in a more psychological sense. Greeno (1974) for example distinguishes two main ways they can contribute to psychological knowledge.

Firstly providing a psychological interpretation of the abstract states postulated by the model can enable it to be used, in the same way as any theory, to explain in meaningful terms the results of experimental manipulations. Or alternatively, the order of inference can be reversed and the psychological interpretation of the components may be provided by examining the way these components are affected by intuitively meaningful variations in experimental conditions. This process has been characterised by Falmagne (1974) as 'second-order experimental psychology', where we infer an underlying psychological process from the behaviour of a model designed to account for variations in behavioural indices, rather than from direct observations of behaviour. It is thus subject to the same uncertainty over the classification of experimental situations. Falmagne points out the danger of circular reasoning in the direction of inference by classifying situations in terms of the predictions of the model.

This debate over the interpretation of models can clearly be extended to the interpretability of solutions of multivariate analysis.

Most of the widely used methods of multivariate analysis are based on some underlying deterministic model. We can make an important distinction between two classes of multivariate techniques on the basis of the way in which this is specified.

In metric analyses the 'behaviour' being modelled is considered as numerical quantities measured at at least an interval level. The formal model, that is the means by which this numerical data are related to a final representation can thus be expressed in terms of direct (usually linear) mathematical functions. However with the exception of a few techniques such as principle components analysis this deterministic model only applies to idealistic error-free data. Most techniques account for only a portion of the numerical estimates, the remainder being designated as 'error'. However the computational methods are designed to give a 'best' answer in a statistical sense regardless of whether this model is satisfied or not. It is thus often impossible to distinguish true random error from error caused by departure from the model.

These models can thus not be regarded in the usual sense as falsifiable quantitative theories. As well as these limitations there is the additional question considered in the previous chapter of whether psychological data can be collected in numerical form of sufficient validity for them to be used as metric quantities.

An innovation introduced in an attempt to circumvent the strong assumptions of metric models has been the use of non-metric scaling methods. These methods might either accept data in numerical form but assume weaker properties than those of the number scale (for example that it represents only an ordering of stimuli) or alternatively operate directly on ordinal constraints implied by the observational units without going through any intermediary scale values. These could therefore logically be denoted 'transformational' and 'direct' non-metric analyses. The first type is more common and the term nonmetric usually implies only this method. However the latter does represent a distinction of some significance as it removes the constraint that the similarities conform to a weak-ordered scale.

The most common type of model, the spatial models, produce some sort of geometric representation of the data. This implies that in addition to being able to define each stimulus as a vector, or a set of values on the individual dimensions we can also define a distance between pairs of stimuli. In multidimensional scaling this representation is to be constructed using empirical data corresponding to these interpoint distances. Judgements of similarity for example can be considered as psychological distances. By the application of a geometric model (normally Euclidean space) to this type of data a dimensional representation of the stimuli can thereby be obtained.

This represents a very appealing model as it appears to offer a means of obtaining a representation of the perceptual process that is unbiased by selective attention producing instructions as are implicit in the use of rating scales.

However that such a meaningful representation can be obtained is by no means automatic. While a solution can always be calculated this will be meaningless unless the assumptions implicit in the model can be verified in the data. Most multivariate procedures, including MDS, have been in common use without the assumptions involved in the representation being explicitly known, let alone tested. The logical foundation for the use of a procedure such as MDS as theoretical model must be provided by the proof of a representation theorem.

For MDS to be regarded then as multidimensional measurement one needs to show, by means of a representation theorem, that the chosen empirical relation satisfies the conditions necessary for a spatial representation.

The foundations for metric multidimensional scaling have been discussed by Pfanzagl (1968) and others. However, as discussed in the previous chapter it is doubtful whether psychological data be realistically regarded as containing metric information. Attempts have been made to provide foundations for multidimensional measurement based on the qualitative relation of order, as have been done for many unidimensional scales.

The first systematic studies by Beals and Krantz (1967) and Beals, Krantz and Tversky (1968) showed that a weak order over stimulus pairs is not sufficient by itself to guarantee a multidimensional representation, but one required additional axioms regarding the nature of the psychological space. The precise conditions required depend on the nature of the representation being sought.

A variety of representation theorems for simultaneous measurement on more than one scale have been developed . under the general term conjoint measurement. Conjoint measurement involves the assumption that a psychological effect of a stimulus, experimental condition etc is produced by the conjoint effect of values on a set of dimensions. The precise nature of this effect determines the particular model to be employed.

Most conjoint measurement models have used the same qualitative order relation as has been used for unidimensional scales. That is, they derive the conditions under which a scale can be derived for the conjoint effects from judgements of the type 'Effect A is at least as great a effect B', and specify the uniqueness of such a scale. For most models this produces an ordered metric scale, which as the number of ordinal constraints increases, approaches a unique linear scale. Efficient procedures for calculating the set of solutions for ordered metric scales are provided by algorithms such as ORDMET (McClelland and Coombs, 1973; Lehner and Noma 1980). These have the advantage that solutions can only be produced if the pairwise orderings obey the conditions required by the measurement model.

Multidimensional scaling can be considered as a combination of one of the most common conjoint measurement models; additive conjoint measurement (Luce and Tukey, 1964, Krantz, 1968 b) and the unidimensional algebraic difference model (Krantz et al., 1971).

That is, the conjoint effect is assumed to be an additive effect of the individual dimension effects. However, each dimensional effect relates in this case to the difference between two stimuli, rather than the value of a single stimulus.

That is, the individual effects are of the form

$|\phi_i(a_i) - \phi_i(a'_i)|$

and must obey the requirements of algebraic difference structures. That is on each dimension there must be a subtractive law such that the effect of that dimension can be represented by differences between stimulus values.

The additive conjoint measurement model specifies that the conjoint effect of these across dimensions is additive. That is relations on the dissimilarity of stimulus pairs must satisfy the condition

$$\delta(a,a') \geqslant \delta(b,b') \text{ iff } \sum_{i=1}^{m} \phi_i(a_i a_i') > \sum_{i=1}^{m} \phi_i(b_i, b_i') \qquad (1)$$

or equivalently, that we can find an increasing function F such that

$$\delta(a,a') = F[\sum_{i=1}^{m} \phi_i(a_i, a_i')] \qquad (2)$$

The most important condition implied by the ACM model is the independence law; the effects of a given dimension must be independent of the effect of any other dimension.

One general class of functions that conform to the additive difference model (ADM) are the Minkowski distance functions, the most common of which being the Euclidean distance.

Using a scaling procedure based on geometric models using such a distance function thus constitutes assuming the ADM as a model or theory of dissimilarity judgements.

Studies by Krantz and Tversky (1971, 1975), Tversky and Krantz (1968, 1970) have suggested that violations of the axioms of this theory do occur in several different sets of stimuli.

Nygren (1979) provides a probabilistic version of the ADM model which enables the degree of significance of the departure from the axioms in a given set of dissimilarity data to be determined.

These tests are not always applicable because they require that the dimension values be known in advance. However, the results of studies using these on artifically constructed stimuli do not provide much encouragement for the idea that the basic perceptual attributes of stimuli enter into such simple geometric models of perception.

These generally negative conclusions of the measurement theoretic evaluation of MDS as a psychological model imply that any simple interpretation of dimensions as basic perceptual attributes is unwarranted.

MDS thus represents the application of a theory whose content is dictated largely by scalability conditions, whose content is not completely verifiable. Without these restrictions for example the form of the similarity function might be quite different. Gregson (1975,1976) reviews a number of models of similarity based on functions which are not admissible as metrics in a spatial representation.

An alternative approach then might be to regard multidimensional scaling, in fact any method of multivariate analysis, simply as data reduction. That is, rather than psychological theories of similarity they could be viewed as methods of organizing, summarizing or displaying data. Beals et. al. (1968) deny the implication that such methods can be used simply as descriptive statistics, pointing out that a critical feature of the method is the minimization of error. If the underlying model is inappropriate then the procedure must necessarily capitalise on noise in the data to obtain the fit, and thus be inconsistent from one data set to another (related) data set. The consistency of fit is thus a critical consideration, not simply the minimisation of error in each case considered in isolation. The theoretical usefulness of any approximation is clearly reduced if there is no way of telling to what extent the approximation is a good one.

PART TWO

THEORETICAL ISSUES IN MULTIDIMENSIONAL SCALING

CHAPTER III

MODELS AND METHODS FOR DISTANCE REPRESENTATION

Metric Multidimensional Scaling

The problem of seeking a spatial representation for a set of distances can be characterized as finding n points whose interpoint distances, as defined by some distance function, match the experimentally obtained estimates of distances between the n objects. Torgerson (1952) developed a solution to this problem that required the assumption that the observed proximity measures represented interval-scaled Euclidean distances. He based his techniques on theorems developed by Young and Householder (1938), which demonstrated that, under general conditions, a matrix of scalar products could be decomposed into a matrix of dimensional co-ordinates.

The scalar products matrix P for n points in a dimensional space is given, by definition, by

$$P = BB' \tag{3}$$

Where B is an nxr matrix ($r<n$) whose elements specify the projections of the n points on r orthogonal dimensions. The Young and Householder procedure for obtaining B is to obtain the nxn matrix V of eigenvectors of P, and the nxn diagonal matrix L of eigenvalues of P. These matrices satisfy the equation

$$P = VLV' \tag{4}$$

and we can thus obtain B by defining $B = VL^{\frac{1}{2}}$. B can be reduced to nxr by eliminating eigenvalues equal to zero (and their corresponding eigenvectors).

Torgerson's procedure consisted of converting the matrix of distance estimates into a matrix of scalar products. It also incorporated a procedure by which the distances could be derived from estimates considered to represent only interval scales by the use of a suitable additive constant. This involved defining an origin at the centroid of all points and deriving equations to compute scalar products for all pairs of vectors from the origin to the stimulus points.

Gower (1966) and McCallum (1974) have drawn attention to the close relationship between this method and principal components analysis and factor analysis.

<u>Non-Metric Multidimensional Scaling</u>

Shepard (1962a, b) proposed a major innovation in these multidimensional scaling methods. He introduced as the central feature the goal of obtaining a monotone relationship between the experimental distance estimates and the distances in the configuration. Shepard did not however develop any logical foundations for determining the degree of satisfactoriness of any given solution. The object was simply to obtain a perfect monotone relation between data distances and a set of 'disparities', chosen to minimise the deviations between disparities and configurational distances. That is we have a two step minimisation procedure.

$$S \overset{m}{=} D \approx \hat{D} = F(X)$$

Where S is a matrix of similarities (or other psychological distance estimates), $\overset{m}{=}$ means perfectly monotonic, $\approx$ is a least-squares approximation, $\hat{D}$ a matrix of disparities, and D a matrix of distance derived by distance function f on a configuration X.

Kruskal proposed a number of indices of fit, denoted stress, which were variations of the general form

$$\text{STRESS}^2(1) = \frac{\sum_{i<j} (d_{ij} - \hat{d}_{ij})^2}{\sum_{i<j} d_{ij}^2} \qquad (6)$$

Kruskal (1964b) outlined the basis of an algorithmic procedure to produce a multidimensional configuration that minimises stress. This procedure, known as MDSCAL, was the first non-metric multidimensional scaling program.

MDSCAL involved two separate minimisation processes. In the first stage it starts with an estimate of X, calculates D. Finds $\hat{D}$ monotonic on S and as near as possible to D, factors $\hat{D}$ to find a new X and so on, continuing iteration until the process converges.

The second stage then followed a method of steepest descent iterative procedure where the configuration was moved small amounts along the gradients defined by the partial derivatives with respect to stress. The first stage might thus be termed quasi-nonmetric, while the second phase is fully non-metric.

Nonmetric multidimensional scaling thus represents the establishment of a homomorphism between properties of the data and the representation algebraic structures, as distinct from the isomorphism implied by metric models, where all properties of the data are to be so represented (Pfanzagl, 1968).

The theoretical advantage that these procedures appear to offer comes from the assumption introduced by Shepard (1962a,b) that judgements of subjective similarity could be considered as a type of proximity which is monotonic on a true distance measure. Similarity judgements could thus be used as input to non-metric scaling procedures in order to identify basic structural features underlying the similarity judgements. These features, for example the dimensions of the derived configuration, could be taken to represent the basic structure of the perception of stimulus attributes independent of the selective attention effects of separate attribute ratings.

While the monotonicity relation represents a weaker and more defensible assumption, the evidence suggests that even this can be considered to be at best only approximately true, and even then only under certain restricted conditions. Gregson (1976) in a comprehensive review of similarity modelling concludes that similarity estimation is best represented by a class of functions not monotonic on any admissible distance measure.

In addition, multidimensional scalings of similarity judgements have been shown to be more context dependent than the perceptual structure hypothesis would imply (see, for example, Green and Carmone, 1971; Gregson, 1972; Gregson and Mitchell, 1974).

Tversky (1977) also gives a critical review of multidimensional scaling analyses of direct similarity judgements and concludes that most evidence suggests that the set of features which forms the basis of a similarity judgement is not invariant during a scaling task, but changes with the context that is established by each specific stimulus combination.

This sort of dependency clearly cannot be incorporated in any procedure that seeks a unique scaling location for each stimuli.

Further critical evidence is provided by studies which have directly tested the measurability conditions imposed by MDS and which have identified significant violations of these in similarity data (e.g. Krantz and Tversky 1975, Tversky and Krantz 1969, 1970). These conclusions demonstrate that the theoretical relevance of the structure obtained by MDS procedures is far from assured. They clearly imply the need for caution in attributing any psychological interpretation to the representation.

The question of the suitability of the data for dimensional representation is clearly one which should be answered as a consequence of attempting to apply the model. As Guttman (1971) observed, " a good technique of data analysis should tend to be self-critical. It should help answer the question as to whether or not its approach ought to be used at all for the given data".

Methods of Similarities-Distances Mappings

Within the linear monotone distinction introduced at the start of this chapter a variety of different methods can be identified for deriving distance information from empirical similarities judgements. Monotonic models can be distinguished into two further categories depending on the nature of the data involved:

(i) transformational methods derive distance estimates from data considered to be in some sense psychological distance measures;

(ii) relational methods match empirical relations in the data with relations defined on distances in the configuration.

Most commonly used procedures are of the transformational type. Programs of the relational type, which match empirical relations in the data with constraints defined on the formal model seem heuristically more appealing, being defined on a completely non-metric basis, as opposed to the metric approximations entailed by quasi-non-metric models. Procedures designed for this type of data have been developed by Holman (1978), de Leuuw (1968), Johnson (1973).

These methods however present obvious problems in terms of the scalability of the data. The two latter techniques in fact accept data as similarity magnitudes, but utilize only information on their pairwise orderings. This is thus equivalent to using relational data with intransitivities removed. Holman's method uses a betweeness relation and simply attempts to minimise the incorrect distance relations in the configuration implied by the empirical relations in the data.

None of these methods, however, has been shown to offer any significant advantage over normal methods to justify these more cumbersome data collection procedures.

Within the transformational procedures a variety of different transformational principles have been proposed, corresponding to different criteria for minimising the error in the data and different measurement constraints to be adhered to. Most of the variations are described in Young (1975) and will not be detailed here.

Some of these methods of regression impose much weaker constraints on the data than others. For example, in Kruskal's monotonic regression principle any monotonic function can be fitted to the data to achieve greatest fit with the derived distances. The form ofthe function that is derived can be observed by an inspection of the plot of data against disparities.

These monotonic regression models have proved very popular in the period since their initial development and the majority of MDS procedures are of this type. This popularity however appears more a function of its apparently parsimonious theoretical basis, rather than their proven superiority. The choice of the type of regression between data and distances in fact involves compensatory considerations to the maximum tolerance principle of the monotonic transformations.

Programs based on this procedure are increasingly prone to problems of local minima, slow convergence and degenerate solutions. The transformational principle developed by Kruskal in particular tends to increase the proportion of tied data (i.e. several response scores may be mapped into one disparity to preserve the monotone relation). This tends to diminish the response information, promote degeneracy and increase dimensionality. Shepard (1974) in reviewing the problem of degeneracy (solutions with all equal distances) suggests imposing additional conditions on the monotonic regression (such as smoothness). Weeks and Bentler (1979) in a comparison of the effectiveness of linear and monotone scaling models are even more doubtful about the utility of weak monotone models. They compared the performance of both models in the conditions where the linear model is or is not satisfied. They found the linear model to perform substantially better when its assumptions were met, and when they were not the linear model applied to ranked data still proved slightly better than the monotonic model. They argue that the linear model's advantages of conceptual simplicity and computational efficiency justify its use in preference to the monotonic model. If an inspection of the plot of data to recovered distances shows evidence of systematic non-linearities then applying the linear model to ranked data can retain these advantages while incorporating the monotonic model's robustness to these distortions, and avoiding the danger of degeneracy.

In some cases the form of the transformation function between data and distances can be predicted in advance, and in this case the prior transformation of the data to linear distance estimates for use in a metric program is recommended in preference to the use of a non-metric model. One of the few situation where there exists any theory which enables one to specify the form of the prototypical regression function is for confusability. In this case both theory and data suggest it should obey an exponential decay function (Shepard 1957, 1958 a,b) Arabie and Soli (1978) provide an example of the use of an exponential transformation to confusion data using the classic and well-analysed Miller and Nicely (1955) vowel sound data. This had previously been analysed by non-metric MDS and clustering programs (Shepard 1974) and metric programs (Carroll and Wish, 1974; Wish and Carroll, 1974).

Prior application of the transformation to linearity resulted in a solution in 4 dimensions which recovered almost as much data as Carroll and Wish's direct analysis in 6 dimensions (both used the INDSCAL program). It thus seems plausible that the additional dimensions for untransformed data may have been needed to provide additional degrees of freedom for the linear model to fit exponential data (Kruskal and Shepard, 1974). For example one of Arabie and Soli's dimensions proved to correspond with two of Carroll and Wish's dimensions, and overall much of the same information was represented. As well as being more parsimonious, interpretation of the four-dimensional space was shown to be consistent with the ordering of stimuli on the dimensions in the object space, the configuration of weights in the 'subject' space (these in fact corresponded to averages over different experimental conditions), and changes in the weights over changes in the physical conditions.

Krantz (1967) and Gregson (1975) have criticized the logical basis for the derivation of the exponential function, arguing that it is not sufficient to establish its invariance over all possible scaling conditions, although it may happen to be a sufficiently close approximation in the case used for most research, i.e. acoustic confusions.

Alternative Distance Functions

The selection of the appropriate distance function for multidimensional scaling is clearly of prime theoretical importance as this constitutes the form of the underlying quantitative model, by which scale values on dimensional components are related to values corresponding to a unidimensional interval level rescaling of the data.

Most investigations have been restricted to the Minkowski metrics, which, as previously mentioned, represent a major restriction on all possible models. This class of functions was considered by Torgerson(1958) in his original formulation. Torgersonsuggested that the psychological interpretation of the Minkowski parameter could relate to the extent to which the different dimensions of the stimuli were perceptually distinct. Torgerson considered that the basic perceptual processes followed the Euclidean metric, but the more cognitive elements were involved the poorer the Euclidean or even any metric representation would be.

Other investigators have interpreted the metric along similar lines, although many regarded the city-block metric as more basic. Attneave (1950), Hyman and Well (1967, 1968) concluded that when the dimensions are obvious or compelling then a city block metric fits best, with the Euclidean metric operating when dimensions were not perceptually distinct. Wiener-Ehrlich (1978) also found some support for the distinction between analyzable and integral stimuli being related to the form of the best fitting metric, although the results were far from clearcut, and were limited by the use of only the city block and Euclidean metric, and by the prior averaging of data across subjects.

This finding that the city block is the best Minkowski metric for stimuli with perceptually distinct dimensions has also been supported in several other studies involving the direct fitting of various metrics using the known values on physical dimensions (e.g. Roskam 1972; Gregson, 1966).

Shepard (1964) made a similar distinction between unitary and analyzable stimulus materials. He concluded that an at least locally Euclidean metric is required for essentially unanalyzable stimuli, and suggests a different metric $1 < r < 2$ may be required for highly analyzable stimuli.

A slightly different interpretation has focussed on the relative weighting on individual dimensions implied by the various metrics. Micko and Fischer (1970) suggested a relative dimensional salience interpretation of the Minkowski parameter. They pointed out that in up to 2 dimensions one can replace Minkowski metrics by city block metrics which allow for continuous variation in the relative weighting given the two dimensions.

A more plausible dimensional weighting explanation however relates the salience of dimensions to the size of the difference between the stimuli on that dimension. Wender (1968) pointed out that a property of the Minkowski parameter is that increasing values of r imply an increasing dominance of the larger component differences over the smaller. (This is denoted the weighting property). Wender proposed that the metric would depend on an explicit use by the subject of the weighting property, which would in turn depend on the difficulty of the task.

If stimuli are easily compared with regard to their underlying characteristics, and if these characteristics are easily combined into an overall judgement of similarity then the subject will tend to weight all distances equally, and hence r will be 1. As the task becomes more difficult the subject will mainly rely on large differences and hence the value of r will increase. Wender quotes empirical evidence to support this in a comparison of the effect of the exposure times of the stimuli - the best fitting component increased with decreased exposure time. He used stimuli with highly distinct components, so that with long exposure times a city-block metric proved most appropriate. However, at shorter times this tended towards the Euclidean metric.

There are practical as well as theoretical considerations involved in the choice of r. The most common technique of identifying the optimum value of r has been to follow the procedure outlined in Kruskal's initial paper (1964 a). This involves deriving a solution for a range of values of the exponent, and plotting the stress values over the range of r. Kruskal quotes an example of such an analysis of some data on judged similarities of colours due to Ekman (1954). This showed a minimum at about 2.5, although this was only slightly lower than the Euclidean value of 2.0.

Shepard (1974) criticises this practice of simply seeking the model with the lowest residual departures from monotonicity. He points out that the stress values have not been shown to be comparable across different values of r. He also points out that possible degenerate solutions, with large numbers of equal interpoint distances, increase as one departs from the Euclidean metric in both directions. This conclusion is borne out by practical attempts to obtain non-Euclidean solutions, which often produce problems of slow convergence and local minima, (Hyman and Well 1967,1968) and are increasingly sensitive to the choice of a starting configuration (Arabie and Boorman 1973, Arabie 1973, 1978 a,b).

Arabie (1978a) comments on the difficulty in determining the acceptability of obtained stress values across different numbers of stimuli, dimensions, Minkowski metrics and so on.

Shepard thus concludes that while the minimum stress criterion may be of some value in showing that the underlying metric is Euclidean or near-Euclidean it is of little value in identifying a precise Minkowskian value. He points out that, especially in the case of two dimensions, it is difficult to distinguish between different values of r. For example, the dominance and city block metrics in two dimensions produce similarity contours differing only by a 45° rotation and change of scale. Koopman and Cooper (1974) demonstrated a similar pairing between intermediate values of r above and below 2, which hold the relation $1/r + 1/r^* = 1$. This is however not as direct as the relation between 1 and ∞ . Koopman and Cooper derive other relationships between metrics above and below 2 which cast strong doubts on the additive versus dominance combination rule attributed to this distinction.

As Shepard notes, several Monte Carlo studies have established that the Euclidean solution is remarkably robust in the face of some departures from the assumed Euclidean metric. Lissitz and Robinson (1977) for example found that Euclidean solutions tended to produce the lowest stress regardless of the actual distance function used to produce the data. Sherman (1972) also found Euclidean solutions do at least as well under most conditions as true r scaling. Shepard (1966) argued that Euclidean scaling can often determine the form of the underlying configuration even if the underlying metric is of the general Minkowski or still more general semi-metric type, provided only that the space is of the isotropic (uniform distance function) type. The results reported in Chapter V, Fraser (1976) and Wiener-Ehrlich (1978) also support this conclusion.

If the space is non-isotropic then standard MDS programs breakdown. Shepard and Carroll (1965) describes a method of parametric mapping which can accommodate distortions of this type. This method abandons the requirement of the triangle inequality retaining only the notion of a continuous underlying co-ordinate space. This is equivalent to a scaling in Riemann space (and thus the underlying psychological space may be a kind of Finsler space). This method is however liable to severe problems of slow convergence and local minima. Several other procedures have been suggested as means of attempting to identify the form of the distance function.

Piesko (1975) derives a procedure for scaling in Riemann space which allows for estimation of the departure from a uniform Euclidean metric in different sections of the space. Peisko quotes an example of such a scaling of Torgerson's (1958) colour similarities data which showed that this was only slightly non-Euclidean, a conclusion also reached by Messick (1956). (Bentler and Weeks (1978) however showed it could probably be more simply considered as slightly three dimensional). This is therefore a procedure which appears to have some promise, although the technique needs to be much more extensively developed.

Lindman and Caeli (1978) also support the use of Riemannian space as a generalization of Euclidean space. They impose the restriction of constant curvature to prevent the problem becoming so general that its solution is trivial.

No logical justification is provided however for this as the appropriate representation space for similarity scaling.

Another attempt to accommodate the unknown form of the distance function is provided by the technique of maximum variance multidimensional scaling (Cunningham and Shepard, 1974). This approach concentrates on deriving a scaling of distances, rather than the dimensional structure underlying them. The procedure attempts to derive a scaling of the data that maximises the variance of the scale value subject only to (a) the strict satisfaction of the metric axioms of a distance, the most salient of which will clearly be the triangle inequality, and (b) an adequate maintenance of monotonicity with the data.

This approach thus represents a primary concern with the form of the psychophysical function relating the data to distance estimates. However, as they suggest, this could be derived in a second stage process by using the distance estimates derived from their program as approximate distances (Euclidean or Minkowskian) for input to a metric MDS program. The existence of such an underlying co-ordinate space in which to embed the distances is however not a necessary condition for the derivation of the distances, nor is the knowledge of the precise analytic form of the distance function.

Cunningham and Shepard describe a Monte Carlo evaluation of this procedure which showed that it was able to recover the precise form of the similarity function far more accurately than standard non-metric MDS programs. With these methods the similarity function emerges as a consequence of the derivation of the dimensional configuration, in the form of the monotonic transformation that is produced between data and distances. With maximum variance MDS the strong dimensional assumptions are delayed to the stage following the determination of the psychophysical function, and may thus prevent distortions in the accuracy of the former showing up in the latter.

Several doubts have however been raised about the utility of this procedure. Rosenberg (personal communication) comments that it is difficult to see what is hidden behind the maximum variance goal. It may for example force many triangle equalities. While it imposes a weak restriction on the data, it really represents an attempt to avoid specifying the model. The lack of a dimensional representation for example removes one of the main bases for a substantive interpretation of the scaling solution.

Thus in conclusion, the attempts to generalise the concept of distance away from the standard Euclidian metric have not been all that successful. Nor have attempts to use distance functions that are sufficiently general and unspecified to cover most ranges of stimuli and conditions. Even non-metric scaling involves modelling of a very definite nature, and the assumed distance function is an important part of this. The metric space resulting from a scaling procedure may be regarded as a model of the process underlying the similarity judgements only if there is empirical evidence for this assumption.

Combining Data Across Individuals

Most scaling procedures require aggregation of data over individuals to provide sufficient information to derive a solution. A classic problem is how to combine the matrices from a number of subjects so as to not lose sight of any individual differences that might exist. The simplest solution is of course to ignore individual differences and use various averaging procedures. Such average distances will provide more accurate estimates of the dimensional distances provided one can assume that all individuals are relating their judgements to a common subject space. The need to use non-metric transformations between data and distances is reduced with averaged data, as it is easy to see that an average of a series of different monotonic transformations will tend towards linearity. It is thus not surprising that scalings of averaged data tend not to show much difference between metric and non-metric solutions (e.g. Green and Rao, 1972; Green and Carmone, 1970).

Carroll and Chang (1970) describes a program INDSCAL, which uses a weighted Euclidean distance formula, where each subject may apply different weightings to dimensions before calculation of a Euclidean distance.

While this procedure has been criticised as theoretically quite restrictive it does have some intuitively appealing computational features. One such feature that appears to aid interpretation is that is produces a solution that is not subject to arbitrary rotation. Several studies, including one reported here (Fraser 1976) have found that when this program is applied to empirical domains with highly distinct perceptually salient dimensions, the co-ordinate structure is usually represented in terms of those dimensions . INDSCAL has been applied with reported success to auditory tones (Bricker and Pruzansky, 1975) more complex sounds (Howard and Silverman, 1976); the perception of nations (Wish 1970, 1971; Wish, Deutsch and Biener, 1970), interpersonal relations, (Wish, Kaplan and Deutsch, 1973; Wish 1976), passages of prose (LaPorte and Voss, 1979) physical environments (Ward, 1977) artistic style (O'Hare, 1976) perception of stereotypes by children (Shikiar and Coates, 1978), and a variety of sets of lexical items (Fillenbaum and Rapoport, 1971).

Several authors have commented that the weighted Euclidean model is in fact a restricted case of a more general model of three-mode multidimensional scaling derived from three-mode factor analysis (Tucker and Messick, 1963 ; Tucker 1964, 1966, 1972). See for example Bentler and Lee (1978), Takane, Young and de Leuuw (1977), MacCallum (1976). Carroll and Chang (1970a) also mention this point but do not accept that the more general formulation must automatically be preferred.

Both the Carroll-Chang and Tucker models can be considered as generalizations of principal components analysis. (Despite the title Tucker's model is not strictly a factor analytic model as it makes no provision for uniqueness. Bloxam (1968) has rewritten the model in a form that includes the concept of uniqueness in the covariance structure).

Tucker and Messick's (1963) model gives a method of identifying homogenous clusters of individuals for whom a separate scaling can be performed. The assignment to clusters is based on the distributional characteristics of component scores. The centroids of each cluster then constitute an 'idealized subject', and are used to calculate scalar products and thus a dimensional scaling from that 'point-of-view'. The dimensions of these idealized individuals' spaces are used to aid the interpretation of the originally derived person components space.

This procedure does not however specify the nature of the differences between the multidimensional scaling spaces for each individual. The three-mode factor analysis developed by Tucker (1966) does make this more explicit. This is a version of components analysis which yields a model which involves both a person space and a common object space for which weights given to the dimensions and angles between the dimensions are functions of the person parameters in the person space. That is, it postulates that an observed three-way data observation can be decomposed into components attributable to each of the three modes, as well as an internal core matrix representing the relations between them. This model can be applied to a variety of types of input data. We are concerned here only with the multidimensional scaling specialization which is based on symmetric interpoint distances data to derive the scalar products; rather than the factor analysis case, which uses profile data as initial data. Tucker (1972) has developed this multidimensional scaling case more fully, and this is the version usually referred to as three-mode multidimensional scaling.

In this case the basic model can be expressed as a scaling of P objects in r dimensions such that

$$B_i = AC_iA' \qquad i=1,N \tag{7}$$

where B_i is a p x p scalar products matrix for individual i, A is a p x r common configuration matrix and C_i is positive semi definite. The diagonal elements of C_i provide for differentially weighted Euclidean distance in a common space i.e.

$$X_i=AW_i \quad , \quad W_i=\mathrm{Diag}[C_i] \tag{8}$$

Where X_i is the space of individual $_i$ from which simple Euclidean distances are generated.

Horan (1969) showed that the latent roots of the matrix of scalar products of the arithmetic means of the A_i matrices show no consistent relationship with the columns of M, in fact overestimating all differences differing on more than one dimension. He showed that the root mean square does not cause this discrepancy.

However many studies have suggested that this assumption of homogeneity of subjects may be unwarranted. Some analyses of combined matrices have resulted in a stimulus configuration which is clearly not representative of any single subject (Silver, Landis and Messick, 1966). A preferable alternative is thus to utilise models which directly incorporate allowance for individual differences.

Most of these approaches have attempted to isolate information unique to each individual from some information that is common to all individuals. For example one of the most common models is the weighted Euclidean model. This has the general form:

$$d_{jk}^{(i)} = \left[\sum_{t=1}^{r} w_{it} (x_{jt} - x_{kt})^2 \right]^{\frac{1}{2}} \qquad (9)$$

The most common method of implementing this model has been the popular INDSCAL program developed by Carroll and Chang (1970a). This method is formally an n-way generalization of Eckart and Young's (1936) two way canonical decomposition of the scalar products matrix. This therefore involves definition of an origin to calculate scalar products, equivalent to specifying the object space configuration common to all individuals, and C_i is a pxp symmetric matrix specifying the nature of individual i's perception of the object space dimensions. Diagonal elements of C_i correspond to weights applied to the object space dimensions by individual i, while the off-diagonal elements relate to perceived relationship among the object space dimensions. The INDSCAL model is thus clearly seen as the restricted case where C_i is diagonal for all individuals, indicating that the object space is orthogonal for all individuals. Thus while being much more general Tuckers model is clearly much less informative, possessing far weaker uniqueness properties.

Despite its restrictive assumptions INDSCAL'S success in being able to display the structure in data from a wide variety of context has led to a popularity such that it might appear to have completely superceded other three-mode approaches. The advantage of no longer requiring a separate solution for each pool of homogenous subjects has been cited as a critical feature (Carroll and Wish, 1974). Other advantages claimed over point-of-view analysis are a more parsimonious solution (Carroll and Wish, 1974) and a model which can be expressed in terms of a linear function of inter-object distances (Carroll and Chang, 1970a).

However several more recent evaluations have taken issue with the conclusion that INDSCAL provides the most satisfactory model of three mode scaling, and have suggested that its popularity reflects more its easy straightforward application, removing complicated decisions regarding identification of homogeneity of subjects or rotation of dimensions, rather than any claim to theoretical supremacy. It does have one practical disadvantage in that it is often extremely slow in producing convergence and although some later versions have improved this (Pruzansky, 1975) it is much slower than comparable methods.

Another source of criticism of INDSCAL has been on the grounds that it does not constitute an ideal solution to the model it claims to represent. This model has been implemented in slightly different form in quite a number of different procedures. The model itself has in fact sometimes been credited as being independently suggested in three sources published at around the same time; Horan (1969), Bloxam (1968), and Carroll and Chang (1970a). However Horan's paper was in fact originally submitted well in advance of the others (1964), being published posthumously. Also, as pointed out by Takane, Young and DeLeeuw (1977), the Carroll and Chang model does in fact possess weaker uniqueness properties than Horan's original formulation.

Horan's basic model can be most generally expressed as the idea of subjective metrics in a common space. These can be expressed as a set of subject-specific diagonal weight-matrices D_i^2 such that, if A is the pxm matrix of p stimuli in m-space then the subjective metric distances are given by

$$d^2_{jk.i} = (\alpha_j - \alpha_k)' D_i^2 (\alpha_j - \alpha_k) \qquad (10)$$

where α_j is the j^{th}p co-ordinate vector from A. Or, equivalently, in scalar produce form,

$$B_i = AD_i^2A' \qquad i = 1,N \qquad (11)$$

where $B_i = (b_{jk.i})$ denotes the i^{th} subjects scalar product matrix.

The arbitrary origin to define both the coordinate vectors and scalar products is set at the centroid of the configuration. This means that the scalar product matrices are doubly-centered, with their sums of squares proportional to the sample variance of the scalar products.

A feature of this model which have been heavily emphasised by Carroll and Chang in promoting the INDSCAL version of this model is the orientational invariance of the co-ordinate system In constrast to most similar matrix decompositions (as in factor analysis) rotations of the axes are not admissible in this model as they would destroy the diagonality of the subjective metrics.

Both the INDSCAL model and the Tucker model can be seen as generalised versions of Horan's model. Tucker's model (Eqn 7) removes the restriction for D_i to be diagonal.

The INDSCAL model incorporates additional weighting factors used to normalise the solution so that all N matrices have equal variances. The INDSCAL model can be expressed as:

$$B_i = c_i AD^2 A' \quad , \quad tr(AD_i A')^2/p^2 = 1 \quad , \; i = 1,N \qquad (12)$$

The c_i are the N additional unknown used to normalise the Bi . However, as Takane et al (1977) point out, the consequences of this tacitly matrix-conditional procedure have not in general been recognized. They correctly observe that this prevents comparisions of weight across subjects on each dimension as this invariance has been lost by the normalisation, thus destroying one of the very objects of the analysis. Takane et al in fact incorrectly imply that Horan's unconditional model allows both this and comparisons of weight ratios across dimensions, which is not permissible as the weights are on different ratio scales across dimensions (Schönemann, James and Carter, 1977, 1979). This implicit normalisation in the INDSCAL procedure also implies that little significance can be attached to the spread of dimension weights across dimensions (or across studies).

Since subject difference information is often utilised by investigators in providing a substantive interpretation of the stimulus space, as well as vice versa, the limited interpretability of this information is potentially quite misleading (see, for example, Offenbach, 1979). MacCallum (1977) also showed in a Monte Carlo study of the effect of the matrix conditionality assumption of INDSCAL that comparisons across subjective weightings are meaningless unless normalisation is done across the entire data set. Schöneman et al draw attention to the need for certain conditions to be satisfied before Horan's model can be appropriately applied. They argue for the need to test explicity the two critical defining assumptions of the model; the common space condition and the diagonality condition. Such tests have for example been incorporated in their subjective metric analysis program COSPA (for Common Space Analysis) (Schönemann, Carter and James, 1976).

This algorithm is in fact an algebraic solution of Horan's model derived by Schönemann (1971, 1972). This solution is in itself of only limited significance for, as Schönemann admits, algebraic solutions have a tendency to break down with fallible data. It does however provide a means of obtaining an initial configuration which could be utilised by other more statistical procedures, and is used for this by Takane et al's ALSCAL procedure.

In addition, Schönemann's solution can be criticized as involving a number of quite arbitrary assumptions (de Leuuw and Pruzansky, 1978; Lingoes and Borg, 1978b). However this has provided an alternative approach to the construction of algorithms to solve this problem. de Leuuw and Pruzansky (1978) describe an algorithm SUMSCAL which is based on an analytic approach derived from Schönemann's solution. This has so far proved to perform extremely similarly to INDSCAL but is at least ten times faster. The loss function is redefined to allow for noisy data - that is the analytic solution is assumed to be interpreted in terms of a 'fuzzy' set of equations.

Tzeng and Landis (1978) have criticized the preference of the weighted Euclidean model over the more general model of points-of-view analysis. They claim that many of the difficulties associated with the Tucker model can be eliminated. They criticise the assumption that INDSCAL dimensions correspond to 'fundamental' perceptual or conceptual processes i.e. that they are the basic implicit criteria employed by subjects in discriminating among different objects (Carroll and Wish, 1974). They make the valid point that the statistical uniqueness of the dimensions does not guarantee that they represent underlying psychological dimensions, although the reasons advanced for this seem somewhat misconstrued. Much the same point has also been made previously by Schönemann (1972). Schönemann feels that the attachment of psychological significance on this premise alone 'stretches the point a bit', and compares it with principle components analysis where the fact that components are unique does not guarantee they will be psychologically meaningful or useful.

Tzeng and Landis develop what they describe as a composite model of INDSCAL and three-mode scaling, which avoids the unnecessarily strong assumptions of INDSCAL and the problems associated with the Tucker model. This is however basically a points-of-view approach with some alterations to the clustering procedures.

A more significant theoretical limitation of the commonly used methods of three-mode scaling relates to their ability to identify the nature of the differences between individual configurations. The INDSCAL model adopts a very simple relationship between individual and common spaces. Lingoes and Borg (1978b) have criticized the choice of the loss function which is defined on scalar products, as being inappropriate to detect departures from the dimensional salience model. Scalar products are interpretationally less direct than distances. A loss function defined on them thus does not represent a sound statistical basis for identifying whether individual subject differences represent admissible transformations or meaningful substantive distortions of the group space. That is, individual differences of a more complicated nature, involving for example changes in the relative location of points, or non-linear distortions of the configuration, cannot be compared with a fitting of dimensional weights on any sound statistical basis.

MacCallum (1976a) provides some empirical confirmation of this point in a Monte Carlo study of the effect of departures from the assumption of orthogonality of the object space dimensions. He found the index of fit remained surprisingly high regardless of the actual relationship between the dimensions. However the actual configurations produced under non-orthogonal conditions proved to be a significant transformation of the underlying object space. The derived dimensions, no longer corresponded in any obvious way with the actual dimensions, although the basic structure of the data was retained. The derived dimensions could be related to the true dimensions by oblique rotation, but since rotation of any sort is not admissible in the INDSCAL model this is not an indication of successful reproduction.

This shows that the index of fit is not very sensitive to non-orthogonality of dimensional perception and the distortion produced by this, and thus this type of index cannot be regarded as an index of the appropriateness of the INDSCAL model.

Lingoes and Borg (1978) argue for a more comparable testing of competing models to explain the nature of individual differences. They maintain that any such approach should possess the following key features:

(i) the loss function should be designed to represent directly the adequacy of the model, and should be comparable across competing models,

(ii) the fit of the model should be judged relative to the number of free parameters involved in the model.

As a solution to the first point they propose that the index of fit should measure the similarity between the individual and group configurations implied by the model, rather than using scalar products or interpoint distances. The index of fit is the squared product-moment correlation between corresponding elements (Lingoes and Schönemann, 1974).

The set of competing models is classified in terms of the type of transformations of a common objects' space that are permissible to achieve a fit to the unique individual spaces. The index of fit is calculated after the individual spaces have been optimally transformed under the standard admissible transformations that leave distance ratios invariant. Denoted similarity transformations these include rotations, reflections, translations and central dilations. The inadmissible transformation allowed by the model are applied to the common space, to maximise the fit with the optimally transformed individual configuration.

Lingoes and Borg developed a series of five models indicating five distinct classes of relationships between individual and group spaces. These include the standard average subject space, where similarity transformations only are allowed, the dimensional salience model (individual dimension weights are allowed), and the dimensional salience model with idiosyncratic orientations (individual rotation of the group space before weighting).

The dimensional salience model has been previously defined (Equations 11 and 12). The idiosyncratic orientations model involves a generalization of Equation 11 to allow for nondiagonal D_i^2. That is, we have

$$B_i = AC_iA' \qquad i = 1, N \tag{13}$$

as in Equation 7. However a different decomposition of the C_i is proposed in this case compared to the Tucker model. In this case we assume a decomposition

$$C_i = T_iD_i^2T_i' \tag{14}$$

where $T_iT_i' = T_i'T_i = I$ and D_i is diagonal. Thus geometrically this corresponds to an orthogonal idiosyncratic rotation of A by T_i before idiosyncratic dimension weighting with D_i.

All of these models correspond to previously developed procedures. The first is represented by a variety of nonmetric scaling progams which can accommodate separate monotonic transformations across replicated data (e.g. POLYCON, replicated ordinal ALSCAL). The second is represented by INDSCAL, and the third by a generalization of INDSCAL called IDIOSCAL (Carroll and Chang, 1972). The idiosyncratic orientations are referred to as idiosyncratic frames of reference in the IDIOSCAL terminology. However, as previously commented, this represents an overly optimistic view of the extent to which the INDSCAL (or IDIOSCAL) procedure can recover dimensions which reflect underlying psychological processes.

Lingoes and Borg also describe a new class of model denoted perspective models. These involve a generalization of Equation 8 to allow weights of the form

$$X_i = V_i Z \tag{15}$$

i.e. the common configuration Z is transformed by a diagonal matrix V_i operating on points rather than dimensions. Thus this model allows the vector corresponding to each point in Z to be multiplied by a separate scalar - that is the vector corresponding to each point can be stretched or shrunk. A more general version of the perspective model also allows for an idiosyncratic translation of the origin of the common space to further improve the fit. (While normally admissible this is defined as inadmissible in this case). In terms of the psychological relevance of these models they are denoted as individual perspective with fixed origin, and individual perspective with idiosyncratic origins models respectively.

The dimensional salience and perspective models strictly speaking represent two distinct hierarchies in the sense that each model is a special case of the next model up the hierarchy, and also in terms of increasing numbers of free parameters. However, within the usual empirical conditions where the number of objects is much greater than the number of dimensions they can be considered as a single sequence, with the most restricted perspective model involving more parameters than the most general dimensional salience model. The perspective models thus represent a very general class of model, and for stimulus set sizes in the normal range allow a substantially higher dgree of freedom in fitting each individual point into the group space. Such generality is of course bought only at a cost of substantially reduced uniqueness, and it appears that these models are too general to convey much useful information.

However a full range of models ranging from the most restricted to the most general is clearly advantageous for testing various hypotheses regarding the precise degree of common perception for a given set of three-way data.

Lingoes and Borg present some examples of the use of a program denoted PINDIS (Procrustean Individual Differences Scaling: Lingoes and Borg 1976 a,b, 1977 a, Borg and Lingoes 1976, 1978, 1979, Borg 1977) which compares the progressive improvement in the fit of the five models in the sequence above. As they indicate, clearly if the improvement in the fit is not significant in relation to the increased degrees of freedom in the model then little significance should be attached to that class of inadmissible transformation.

This provides a means of comparative evaluation not available in the use of a single procedure. For example in using models from the IDIOSCAL family there is no sound basis to evaluate the significance of the individual differences in the model. As Borg and Lingoes (1978) have observed the degree of scatter of the weightings in a subject space is not a good indication of the amount of variance accounted for by individual weighting effects.

Table 1 shows the individual transformations within which the best fitting solution can be calculated, and the number of inadmissible fitting parameters (i.e. those to which some psychological significance should be attached) for each of the five models in the hierarchy.

The simplest case, involving equally weighted individual matrices, with similarity transformations only, is measured by the squared correlation between individually scaled configurations X_i and the optimal common configuration Z, allowing only transformations of X_i which preserve distance ratios (i.e. rigid motions and central dilations). This optimally transformed X_i is denoted $\tilde{X}_i$, and the fit index is thus $r^2(\tilde{X}_i, Z)$.

The subjective metrics models are then fitted by variously defined transformations of Z. The dimension weights matrix W_1 and the vector weights matrix V_1 represent the inadmissible parameters related to dimensional or directional salience. The parameters related to idiosyncratic orientation are represented by individually optimized rotations Z_i^r or translations Z_i^t, rather than compromise optimisations Z^r, Z^t for all X_1.

TABLE 1

Summary of PINDIS Transformations

Model	Number of Inadmissible Fitting Parameters	Fit Index
Similarity transformation (unit weighting)	0	$r^2(\tilde{X}_i, Z)$
Dimensional salience (dimension weighting)	m	$r^2(\tilde{X}_i, Z^r W_i)$
Dimensional salience with idiosyncratic orientation	$m+\binom{m}{2}$	$r^2(\tilde{X}_i, Z_i^r W_i^r)$
Perspective model with fixed origin (vector weighting)	n	$r^2(\tilde{X}_i, V_i Z^t)$
Perspective model with idiosyncratic origin	n+m	$r^2(\tilde{X}_i, V_i^t Z_i^t)$

From Lingoes & Borg (1978, p.494)

A variety of examples illustrating the use of PINDIS over different types of data sets have been provided by Lingoes and Borg. Lingoes and Borg (1978) reanalysed some data from a previously published study by Feger (1974) using similarity ratings of German political parties. Very little additional variance was explained by the weighted Euclidean (INDSCAL) model compared with the unweighted case (less than a 1% improvement). The perspective model on the other hand improved the fit from an average of 77% of the variance to 93%.

This study indicated that while most individuals, apart from one outlier, were nicely distributed around the centroid, the group space did not seem to represent any individual particularly well. The one outlier individual showed quite an idiosyncratic scrambled configuration which was poorly fitted by the dimensional salience models but which was successfully 'unscrambled' by the perspective model.

A similar situation emerges in most of the other examples discussed. For example Borg and Lingoes (1978) reanalysed the breakfast cereal data analysed extensively in Green and Rao's book <u>Applied Multidimensional Scaling</u> (1972). They found that 72% of the variance in the individual configuration could be explained by the similarity transformations alone, with only an additional 2½% when individual dimension weightings are allowed, implying that no real differential information is involved. This of course negates the considerable efforts made by Green and Rao to interpret clusters in the subjective weightings space.

In contrast, Lingoes and Borg (1978) describe the use of PINDIS to demonstrate the relationship of colour deficient and normal subjects in a study previously analysed by Helm (1964), Helm and Tucker (1962) and others. While no idiosyncratic transformations are required for the normals the colour deficients showed substantially improved fits (15%) with dimension weightings, reflecting the greatly reduced salience of their deficient continuum.

A rather different application is provided by Lingoes (1977c) in relating the similarities among species of fish by geometric transformations. While somewhat esoteric this in fact serves as a useful pictorial analogy of the transformations involved in the series of models.

The variations in the structural models considered so far do not exhaust the possible varieties of individual differences. Isaac (1968) noted that there were two distinct hypotheses to account for individual differences.

(1) The response bias hypothesis - differences in response habits, preferences and strategies account for differences between people in their responses.

(2) The perceptual structure hypothesis - people differ from one another in their responses as a result of differences in the perceived structure of a stimulus set.

Option (1) has not been developed to any significant degree in metric MDS, apart from procedures to remove biases prior to input to a scaling program by the application of a model such as Luce's (1959) model of choice.

Non-metric programs however can provide a means by which these effects can be accommodated by allowing each individual to have a separate monotonic transformation between data and distances. This procedure was developed by Carroll and Chang (1974) as a non-metric generalisation of the INDSCAL individual differences scaling model, and by Young (1973) as an individual differences generalisation of the non-metric models of the M-D-SCAL/TORSCA series (Young and Torgerson 1968; Kruskal, Young and Seery, 1973).

Carroll and Chang's version uses a two stage minimisation, the first of which minimises the metric index of fit to the scalar products (denoted STRAIN) while the second minimises the non-metric index STRESS (Kruskal, 1964a) which is defined on the raw data. As has been observed (Takane, Young and de Leuuw, 1977), there is therefore no logical assurance of convergence to a stable point.

Young's (1973) procedure did not provide a separate solution for the subject weights. The generalisation of this line of algorithm development in a procedure known as ALSCAL (Alternating Least Squares Scaling) is described in Takane, Young and de Leuuw (1977).

The major innovation introduced by Takane et al's procedure is in the nature of the minimisation process. The alternating least squares principle (ALS) described in Young, de Leuuw and Takane (1976) represents a major improvement in this aspect of algorithm construction. ALS is in fact an example of the principle of optimal scaling, proposed by Fisher (1946) as an ideal for scaling subject to constraints, but rarely achieved in practice. Optimal scaling is a general principle for parameter estimation when the parameters must be considered as a number of different subsets. The principle requires that least squares estimates be obtained for one subset assuming the others to be constant, repeating this estimation process progressively through all the subsets, and then iterating the whole process until convergence.

Such a subdivision is clearly implied by Kruskal's monotonic estimation procedure. This procedure, used by most non-metric programs, involves a two stage alternation between satisfying the ordinal constraints in the data by operating on disparities, and minimising the error (stress) between data and the model being applied by operating on distances in the configuration.

An optimal scaling reconceptualisation of this process would require that the observations be rescaled so that (a) they fit the model as well as possible in a least squares sense, and (b) the measurement characteristics of the observations would be strictly maintained. Clearly the disparities satisfy (b) under an ordinal model, but rarely is (a) completely satisfied. Most procedures can only move some predetermined distance along a steepest descent path to minimise departures from fit, but do not guarantee a complete minimum. Thus ALSCAL, by finding the exact minimum, represents a major improvement over previous algorithms, and this is reflected by its greatly increased efficiency, usually reaching its stopping criterion within 4 or 5 iterations. Lingoes and Borg's PINDIS procedure includes an optional ALS process but the solution is not available for all options of the program.

The ALSCAL procedure was also designed to incorporate a significantly greater degree of flexibility in the choice of the appropriate measurement model.

The view of data developed by Takane et al is a somewhat simplified behavioural categorization, and is not really compatible with the fundamental measurement perspective described earlier. They regard all data as basically categorical, and distinguish classes of data according to three types of restrictions; denoted process, level and conditionality restrictions. Process restrictions relate to relations among objects within a category, level restrictions concern the relationships among all the observations between different categories, and conditionality restrictions concern the possibility of distinct sets of categories. Process restrictions primarily reflect the distinction between discrete and continuous underlying processes, and thus reflect relationships that must hold between distances corresponding to observations placed in the same category. Level restrictions similarly relate to constraints that must be satisfied by the distances corresponding to observations placed in different categories. Conditionality restrictions relate to a partitioning of observations into sets that can be meaningfully compared with each other.

The ALSCAL model can be used in a simple or weighted Euclidean model, as well as a range of other distance functions.

It can also be used with a variety of degrees of strength of assumptions regarding the level of measurement of the data (although the data must be assumed to be originally scaled to a particular level, rather than being in basic pairwise relations).

The ALSCAL model can thus allow for individual differences in the response process, in the judgement process, or both (assuming individual differences in the judgement process can be represented by weights in the weighted Euclidean (or other) model).

This model thus represents a significant improvement over the INDSCAL model, both in flexibility and efficiency, and is clearly the best individual difference model currently available. It can however be subject to some of the same criticisms in that it represents the application of Horan's subjective metrics model without any means of assessing whether the model is appropriate for the data.

Alternative Data Types for Distance Scaling

A topic of some importance in terms of the theoretical status of multidimensional configurations derived from similarities or other distance related data is the form of the empirical observations involved in collecting the data. As discussed in Chapter II the information derived from pairwise orderings, complete ranking, magnitude estimates etc., have quite distinct theoretical rationales.

The data used as input for MDS procedures have been overwhelmingly dominated by the direct magnitude estimation of pairwise similarities. If these MDS solutions are to be regarded in any sense as a psychological model then this will clearly need to be from a perspective that recognizes magnitude estimates as basic data. Some studies have used alternative data collection methods, in most cases in order to compare the representations produced by different methods.

In some cases the main object was to alter the nature of the cognitive judgement in order to investigate the stability of the configuration. As such they are primarily directed to establishing the psychological validity of the interpretation of the results rather than the collection methods themselves, and thus will be reviewed in the next chapter. The present section will review the various data collection methods, the procedures required to transform data into a distance matrix format, and studies relating to the comparability of these different procedures.

Second Order Comparisons

While the paired comparison presentation method has been by far the most common, other levels of comparison have also been used. By level of comparison we mean whether judgements have to be made between stimuli directly (as in paired comparisons) or between the outcomes of such a first-order comparison (as in for example, ratio estimation, cross modality matching, etc.). See Gregson (1975) for a discussion of this point. Second order comparisons raise additional problems regarding the consistency of these judgements. With dominance data a frequent source of error is the circular triad, which for true scores would indicate intransitivity in the scaling dimension. For distance data there is no such measure of inconsistency on the orderings of triples, although for magnitude estimation an unaxiomatic error is represented by the triangle inequality, i.e., the direct similarity between two points must never be greater than the sum of the similarities between them and an intermediate point. As Tversky (1977) points out it is easy to construct plausible violations of this axiom, although the actual incidence would undoubtedly depend on the type of data.

In terms of the ordering of similarities one can construct logical violations on quadruples of similarities. Clearly the proximity of three stimuli to a fourth should obey dominance orderings, and are thus contradicted by circular triads. That is, we should have:-

$$S(A, X) > S(B, X) \text{ and } S(B, X) > S(C, X) \Rightarrow S(A, X) > S(C, X)$$

This logical inconsistency is rarely considered with magnitude estimation similarity data, but it becomes an inescapable problem if the similarity data are collected by orderings on this level of comparison. That is, if subjects are asked to order the similarities of two pairs of stimuli (or more usually to select the closest pair from a triad) then such ordering should be consistent with the relation stated above.

One solution is to collect data in a form which forces the subject to eliminate these effects. Multiple rank order data, where the subject orders a set of stimuli in terms of their similarity with a standard provides, an example of such a procedure.

Most studies on the comparabilities of these methods and their respective consistencies have been done using preference data. However, there is a clear parallel between this and similarity data as indicated by the relation above.

Rounds, Miller and Dawis (1978) compared multiple rank order and paired comparison methods as applied to preference data. They found a very high correspondence between the orders implied by the two methods, and equivalent test retest stabilities.

Hendel (1977) showed that the rate of intransitivities in preference orders tended to be constant over all stimuli in a given set, but showed marked differences over different stimulus domains and groups of subjects.

Roskam (1970) suggested the method of triads may present advantages over both ranking and pairwise magnitude estimation procedures in that it involves relations, not magnitudes on similarities, but without large simultaneous presentations.

In the method of triads the subject is typically instructed to select the two most similar in a set of three stimuli (thus ordering all three pairs). Roskam compared several methods of obtaining a rank-order of similarity of all pairs of stimuli from triadic data.

The TRICON procedure does not admit of a unique solution because information is not available on pairs of pairs without a common element.

Roskam demonstrated that the vote count procedure fails to reproduce known similarity orderings from triads. A Monte Carlo study, based on the 15 stimuli for incomplete triadic data (i.e., based on a study reported by Levelt, Van de Geer and Plomp, 1966) showed the correspondence between true rank order and vote count as expressed by Kendall's Tau to be $.72 \pm .03$, indicating that the data used in this study may well have been different from the true rank order.

Roskam suggests it would be preferable to obtain a conditional rank order from triadic data, and use this in a program designed to accept this type of data. However, even this fails if intransitivities occur. In this case the ideal option would be to adapt the scaling algorithm to operate directly on the triadic data.

He presents evidence from a Monte Carlo study showing that this performs better than both rating and ranking data in conditions of high error.

Ratings and ranking methods are more appropriate to the low error conditions, but do not perform as well under conditions of high error.

Complete Presentations A common extension of second-order comparisons involves the case of multiple or even complete presentations, where the subject must make judgements on the outcomes of a large number of first-order comparisons. Complete presentation methods minimise the problem of inconsistency of judgements, although one clearly loses all control over the systematic nature of the decision process. One cannot be sure for example, that the subject has evaluated all alternatives before making his selection from a complete set. The precise nature of the task the subject is required to perform in complete presentation designs can vary to some degree, but will usually fall into one of the two broad categories of sorting or ranking procedures.

Sorting Procedures This is a commonly used class of data collection procedures especially in situations involving large stimuli sets. Rao and Katz (1971) have further classified sorting procedures into three main classes; subjective grouping, pick and order methods.

Subjective Grouping (Clustering) In subjective grouping methods the subject is presented with either the entire or a subset of the stimulus set (usually on cards) and is instructed to sort them into piles according to their similarity. Variations may be to have K, the number of groups, fixed or variable (with an upper limit K_o fixed) or to ask for a hierarchical ordering of the groups. Alternatively subjects may be asked to use a multiple sorting procedure, where they can repeat the sorting several times and thus allow for more than one sorting strategy.

Pick Methods Pick methods also involve grouping similar stimuli, but in this case the subject is presented with a standard and has to pick K group members out of the remaining n - 1. Again K may be fixed or variable. This is sometimes called clustering about a nucleus.

Order Methods Order methods involve the same procedure as pick methods except with the additional requirement that the K stimuli be ordered with respect to their similarity with the standard. Again K may be fixed or variable. (This thus represents a hybrid sorting and ordering procedure).

Rosenberg and Kim (1975) compared the results of single and multiple sorting strategies as applied to Kinship terms.

The significant differences between the two procedures were measured in terms of the INDSCAL configuration obtained from a scaling of the combined data. For example an obvious dimension, sex, was rarely used when only one sorting was permissible. This illustrates a difficulty with these methods in that each subject is clearly restricted to using a fairly simple sorting strategy, and one must depend on individual differences to produce a representative coverage of the salient features in the stimulus domain.

Such a space could not therefore be necessarily expected to model the psychological processes involved when a subject is free to combine all the salient features he wishes to utilise to produce a numerical estimate of similarity magnitude.

However, sorting procedures have been successfully used for multidimensional scaling of such stimulus domains as occupation names (Burton 1972) nations (Wish, Deutsch and Biener 1972) and ethnic groups (Jones and Ashmore 1973).

Distance Formulae for Sorting Procedures These procedures clearly require a means of deriving a distance measure from the subjective groupings before they can be used as input to MDS procedure. This must be based on the assumption that more similar objects will more often be sorted together, although a wide variety of indices have been suggested to quantify this in terms of a distance estimate. Clearly the form of the index will vary according to whether a grouping or pick index is involved, as the former involves a partitioning of the set of stimuli while the latter, by replacing stimuli, allows for overlapping groups.

Rosenberg and Kim (1975) suggest the formula

$$\delta_{ij} = \sum_{k=1}^{K} (s_{ik} - s_{jk})^2$$

where s_{ik} is the frequency of sorting i and k into different groupings. For a single individual S_{ij} would thus be two valued, say '0' for the same group, '1' for a different group. For pooled data s_{ij} would range from 0 to N, the number of individuals. The formula is thus a sort of squared Euclidean distance on the set of these values over all other items.

This procedure is thus equivalent to treating the data as profiles, the components being the other stimuli, and computing Euclidean distances between them. Two major limitations are apparent with this procedure.

It is not sensitive to differing sizes of groupings, nor to the possibility of inconsistencies in pick procedures when different stimuli serve as the standard. This latter problem derives from the factthat the n x n matrix is in fact row conditional. Rosenberg and Kim's procedure deals with this simply by averaging across the two triangular half-matrices, a procedure that does not appear too suitable to preserving the constraints implied in the original data. The problem of treating conditional distance matrices will be considered with respect to data derived from ranking procedures i.e. where all stimuli is ranked with respect to their proximity to a standard stimuli, with all stimuli in turn serving as standards. While sorting matrices, can be considered as degenerate forms of this type of data matrix, the distance information retained is rather limited. Pick and order methods can be most reasonably regarded as an incomplete form of complete ranking data with the data relating to the larger (and hence less reliable) distances missing. The more relevant problems with sorting procedures, however, is not to remove conditionality but to extract the maximum information regarding distances.

The most common approach has been to use information-related techniques. The information contained in placing two stimuli in the same group is used as a measure of the distance between them (e.g. Burton; 1972). However, set theoretic (Restle, 1959) and graph theoretic (Flament, 1963) measures have also been suggested.

Arabie and Boorman (1973) criticise this faith in entropy measures to deal with all situations. While the measures have had some axiomatic justification (e.g. Khinchin, 1959) they have not been supported in any systematic comparison with rival structures. Arabie and Boorman tested 12 possible indices at distance between partitions and found a measure PAIRBONDS, which compares observed pairings with the set of all possible pairings, gave consistantly lower stress than information theoretic measures.

This measure was

$$\text{PAIRBOUNDS} = D(P) + D(Q) - 2D(P \cap Q)$$

where

$$D(P) = \sum_i \binom{|P_i|}{2}$$

and

C_i is the cardinality of cell i.

This is of course a measure between the partitionings P and Q, not the stimuli so partitioned, i.e. it is a measure on subjects not stimuli, and thus does not necessarily imply that information theoretic measures are less effective in the latter case. It does however support their view that some empirical justification is needed before they can be assumed to be the most suited for MDS. Arabie and Boorman demonstrated the applicability of their measure in a scaling of the changes in sociometric relations over time in a monkey troup. Thus it is clear that information from sorting procedures can be expected to provide at best limited information relating to the inter-object distances.

Ranking Procedures

Ranking procedures require the subject to generate the complete rank order of the proximities within a row of the n x n stimulus matrix. It thus produces data with no inconsistencies within rows i.e. intransitivities, at the expense of requiring judgements which imply a fairly high level of task difficulty.

If stimulus pairs are comparable on a variety of dimensions, then the task of producing an ordering with respect to a standard stimulus implies that a very large number of initial pairwise judgements both componentwise and overall are to be made. It is thus quite plausible that the subject might adopt a different, more simplified strategy than might be used under pairwise presentation. Inconsistencies in judgements are of course still possible across rows. A method of removing such inconsistencies in converting ranked data into distances was suggested by Coombs (1964) and developed more rigourously into a computer program TRICON by Carmone, Green and Robinson (1968). This program starts with the matrix of ranks assigned to column stimuli with row stimuli as standard. This is converted into a dominance, matrix on pairs of stimuli, with '0's and '1's to indicate whether the row stimulus pair has a higher similarity than the column stimulus pair. This matrix is then powered to give second and higher order effects. The row sums of this powered matrix becomes the distance estimates for the symmetric matrix.

Rao and Katz (1971) applied Monte Carlo tests to compare sorting and ranking methods according to their ability to reproduce a known configuration.

For each method two types of analysis were used - a group scaling (based on pooled data) and a individual differences scaling (an INDSCAL analysis of distances estimates derived from the individual data).

For the subjective grouping methods the group scaling data matrix was simply derived from the number of times each pair was placed in the same group.

The individual distances matrices were calculated by applying Rosenberg and Kim's formula (Equation 16) to the two vectors of S_{ij} values, which in this case would contain only 0's and 1's.

The pick k/n-1 methods were treated in substantially the same manner. Again the individual distance matrices were calculated by treating the matrix of 0's and 1's as profile data. However since the pooled matrix is also row conditional this was also converted to distances in the same manner prior to scaling.

The order methods were also similarly treated except now the individual matrices have ranks, rather than 0's and 1's. However the row conditionality was removed in exactly the same manner, by treating them as profile data and calculating artificial distances.

The main results of Rao and Katz's Monte Carlo study were:-

- INDSCAL individual differences approach did not reproduce the original configuration as well as the group approach.
- Non-metric group scaling was consistently better than group scaling.
- Pick or order were better than subjective grouping methods; there was no consistent superiority of order over pick.

However it seems highly likely that these results are more a function of Rao and Katz's methods of converting the data into distance estimates than any intrinsic superiority of the different methods. Although not mentioned by the authors the poor performance of the individual differences approach seems undoubtedly due to the arbitrariness of the profile method of removing the row conditionality. This would also account for the surprising result that the extra information on rankings contributes nothing to the quality of the solution.

A second aspect of Rao and Katz's methods which is unsatisfactory is the lack of any adequate compensation for different group sizes. Simple frequency counts or any measures whose relative orderings are not invariant after normalising for cluster size are not very satisfactory distance estimates for a clustering procedure.

Large Data Structures

Often the application of MDS to a substantive problem might imply a need for a fairly large set of stimuli in order to ensure fairly comprehensive coverage of all relevant perceptual properties of that domain. The traditional paired comparisons data collection procedure is not very suitable for this situation due to the rapid increase in number of judgements over increasing stimulus set sizes. Complete paired comparisons require $n(n-1)/2$ responses. This second power increase rapidly becomes prohibitive in terms of the number of judgements required from a single subject for only moderately large stimulus sets (e.g. 300 responses for 25 stimuli).

There is of course a certain amount of redundancy in the complete presentations case. Various analytic solutions to the question of the amount of data required in paired comparisons methods have been suggested (Gulliksen, 1956; Schönemann 1970; Kaiser & Serlin, 1978). These suggest that with carefully selected error free data the percentage of data required can be quite small. However, with inexact data a certain amount of overdetermination is clearly desirable to order to minimise the effects of errorful data. Most MDS algorithms are designed to accommodate missing data, and rarely place any procedural limitations on amount or patterning. While there are several incomplete designs if one does not require to retain individual differences (Coombs, 1964) if these are to be preserved the selection of judgements must be more careful.

Spence and Domoney (1974) as a result of a Monte Carlo investigation of the effects of these factors recommend a "degrees of freedom" ratio (i.e. a ratio of data to free parameters of the dimensional configuration) of at least 3.5. They found cyclic designs and random deletion designs to be superior to an overlapping cliques method suggested by Torgerson (1958) (i.e. elimination of overlapping submatrices).

An alternative approach is to use a two stage scaling approach. A base set of n_1 stimuli are scaled from complete paired comparisons with the additional $n_2 (n_1 + n_2 = n)$ stimuli being embedded in this space as a result of pairings with the n_1 base stimuli. This procedure was recommended by Fenker (1972), who noted that the standards must comprehensively exhaust the dimensionality of multidimensional psychological space, although did not suggest any way of ensuring this.

A further refinement to this approach is provided by the use of an interactive scaling mode to select the most suitable n_1 stimuli to use for the initial scaling. Young and Cliff (1972) describe an interactive scaling program for individual subjects scaling. Young and Cliff's principle attempts to optimise the determination of nr scale values (for n stimuli in r dimensions) from less than $n(n - 1)/2$ judgements. From a result by Ross and Cliff (1964) the scale values can be derived from the distances of each point from a subset of $r + 1$ points, provided the latter cannot be embedded in a space of smaller dimensionality. Thus 60 points can be scaled in 3 dimensions from only 230 of the 1760 distances.

This of course requires foreknowledge of the appropriate dimensionality, an appropriate basis, and linear error-free data. It can thus only be used with metric programs under conditions conducive to reliable valid judgements. A similar procedure is described by Kehoe and Reynolds (1972).

Young, Null and Sarle (1978) also describe a method of interactive data collection, rather than interactive scaling. This is a procedure whereby conditional rank order data may be collected by using less than complete presentation methods. Subjects are asked to choose which of a set of comparison stimuli is closest to a standard. The selection of comparison stimuli is however based on previous responses in order to produce the conditional rank order relative to that standard in the smallest possible number of judgements. The comparison list length can be varied from two (method of triads) through to n - 1. Further economies can be achieved by specifying a certain list length within which only a partial order obtains. This thus represents an attempt to obtain the positive features of triadic presentation methods with the regularities of complete presentations .

Indirect Proximity Measures

Many measures can be regarded as being at least to some degree related to the proximity of the stimuli. As well as the direct estimates there are also indirect behavioural measures such as confusion and generalization probabilities. One major difficulty with using these as distance measures is that they are not symmetrical; the probability of confusing i as j is not necessarily equal to the probability of confusing j as i.

We can distinguish two main approaches to this problem:-

i) <u>the direct methods:</u> The assymetric matrix is scaled directly. For example, the basic input could be a matrix L on quadruples of stimuli, with entries +1, 0, -1. A + 1 in cell ℓ_{ijkm}

$$=> S_{ij} > S_{km} => d_{ij} - d_{km} > 0$$

Programs exist that work directly on pairs of distances from information on quadruples of stimuli e.g. CDARD-7 (De Leeuw, 1968), PARSCL (Johnson, 1971).

An alternative type of program uses a type of unfolding analysis to calculate a row or column solution (See following section).

ii) <u>the indirect methods:</u> These involve ways of removing the assymetry from the matrix. Shepard (1957) derived a method of converting confusion probabilities into distances from a complex and extensive model of stimulus and response generalization. Shepard's formula, a sort of transformed geometric mean was:

$$d_{ij} = .5 \log \frac{[f(i/j) \quad f(j/i)]}{[f(i/j) \quad f(j/i)]}$$

where f(j/i) is the frequency that stimulus i is identified as j.

Wilson (1963) suggested a measure based on a more simple assumption that independent differences in cues should have additive effects in reducing the confusion error. So his measure was a transformed arithmetic mean:

$$d_{ij} = \sqrt{1 - \frac{f(i/j) + f(j/i)}{f(i/i) + f(j/j)}}$$

In both cases the transformation was applied because they were developed for metric MDS programs, so besides removing assymetries the measures had also to be linearly related to distances.

For use with nonmetric MDS programs the transformations can be dropped, and so the measures become;

Shepard's

$$S_{ij} = \frac{f(i/j) \quad f(j/i)}{f(i/i) \quad f(j/j)}$$

Wilson's

$$S_{ij} = \frac{f(i/j) + f(j/i)}{f(i/i) + f(j/i)}$$

Most empirical studies on confusions have been done with aural stimuli, usually vowel sounds.

Wilson compared the two formulae in their metric form and found that both distinguished the same three dimensions involving articulate and acoustic cues, but in slightly different manners.

A better comparison of these various techniques was done by van der Kamp and Pols (1971), also using aural stimuli - in their case 11 Dutch vowels. As well as the methods previously described they included one more indirect measure. This was the inclusion of a response bias factor i.e. the actual confusion matrix is assumed to be a product of a symmetric and a response bias matrix e.g. for two stimuli i, j the product of

$$\begin{array}{c c} & \begin{array}{cc} i & j \end{array} \\ \begin{array}{c} i \\ j \end{array} & \begin{bmatrix} A & B \\ B & A \end{bmatrix} \end{array} \quad \begin{bmatrix} 1 - \lambda & \lambda \\ 0 & 1 \end{bmatrix}$$

gives the obtained 2 x 2 matrix. λ is the bias associated with stimuli i, j, and by solving for this one can obtain the theoretical symmetric matrix shown above. This method requires that it be appropriate to split the 11 x 11 matrix into all possible 2 x 2 matrices i.e. the response biases on pairs must be independent of the other stimuli (a direct analogy to Luce's (1959) choice axiom). Previous research has indicated that this assumption does hold when Luces axiom is applied in this manner to confusions data (Wagenaar, 1968).

Van der Kamp and Pols evaluated the various methods of scaling confusion data in terms of their ability to reproduce the perceptual structure obtained from direct similarity estimates.

Each configuration derivedfrom confusion data was rotated to maximum congruence with the similarity-based configuration, and the fit measured by correlations of projections on pairs of matched dimensions. The results showed that procedures involving prior transformations to a symmetric matrix all performed better than the direct analysis of the symmetric data. The best transformation procedure was provided by the response bias correction and a simple averaging technique also showed a better fit than the Shepard generalization model.

Various other behavioural measures have been suggested as indirect proximity measures. Steffre (1972) for example, advanced the proposition that an individual will behave similarily towards things which seem similar to him. This sort of relation ship has clear practical applications in situations where the behavioural measures are a dependent variable one wishes to predict from similarities data, for example, in market research. Steffre compared brand switching behaviour with judged similarity for different makes of cigarettes and toilet soaps, obtaining rank order correlations of .74 and .82 respectively. This hypothesis can be taken a stage further by assuming that the dimensional structure implied by the similarity information also governs consumer behaviour. Steffre quotes some evidence for this in the use of this structure to predict buying patterns after introduction of a new product.

Various other measures have been shown to have some relation with similarity. For example Henley (1969) showed some correspondence between free association responses, proximity in free-recall lists, and assessed smiliarity. However in the absence of any well constructed model demonstrating the precise role of similarity in generating these indirect measures they are of little more than passing interest. Certainly a moderate correlation between direct and indirect measures contributes little to establishing the generality of psychological distances as components of cognitive judgements.

Distance Models of Profile Data

Profile data, that is sets of measures of stimuli on defined dimensions, such as rating scales, can also be considered to represent information on distances. That is, each stimulus profile consists of a vector of scores and can thus be represented as a point in a multidimensional space whose dimensions are defined by the profile attributes. One would of course normally expect to be able to represent stimuli by a vector in a space of much lower dimensionality such that the implied distances between them are substantially preserved. If these were defined in terms of the intercorrelations of vector scores, they would then correspond to scalar products in the geometric representationof factor analysis. However, MDS applications require that some more direct measures of a distance between stimulus profiles be preserved. There are two ways to convert data of this type into interpoint estimates; artificial distances and conjoint scaling.

Artificial Distances

Methods of assessing the similarity among profiles is of course a problem not unique to MDS; it occurs for example in clinical or personality test situations. The most widely used method among psychologists in this area was the correlation - usually termed a Q correlation. However, for MDS purposes the most common is the square root of the squared scale differences. The rationale for its use is that it can be regarded as a Euclidean distance in a space in which the rating scales are orthogonal dimensions.

However the arbitrariness of any method of combination of rating scales into distance estimates is reflected in the large number of alternative measures proposed.

Cronbach and Gleser (1953) in a review of the various approaches that have been used show the conflicting attitudes to the problem. Pearson (1928) for example, proposed that distances should be calculated only after the variates had been standardised. Catell (1949) proposed using a correlation index transformed to approximate a measure of Euclidean distance which he termed a 'coefficient of nearness'. He proposed several variations of this which he saw as a sort of standardised distance, i.e. $-1<r<1$, $E(r) = 0$. For example one such measure is:

$$r_p = \frac{K\Sigma\sigma_j^2 - KD^2}{K\Sigma\sigma_j^2 + KD^2}$$

where K is twice the median χ^2.

Some studies have used more complex procedures, although often on fairly arbitrary grounds. For example, Yoshida (1964 a, b) in a series of analyses of ratings of odours, varied between Euclidean and city block distances, and in one case divided his scales into what appeared to him to be qualitatively different groups, calculated city block distances within groups, and took the RMS across groups. The correlation between this matrix and the directly estimated similarities was not surprisingly only 0.16. Another study using Euclidean distances produced a much more respectable 0.62 correlation between direct and derived dissimilarities, especially considering the inconstancies in that particular modality.

A lot less objection would be made to the Euclidean distance formula if it could be shown that the scales chosen were in a one-to-one relationship with the orthogonal dimensions in which the stimuli are embedded.

Even then one still has to make the assumption that the same combination rule is being used, although evidence is presented elsewhere to suggest that this is not critical, as scaling solutions are relatively insensitive to this (Fraser, 1976).

The Mahalonobis distance formula (Rao; 1948) is given by

$$D^2 = (X_i - X_j)\, C_w\, (X_i - X_j)' \qquad (17)$$

Where C_w is the covariance matrix between varieties within groups.

For independent orthogonal factors this becomes equal to D.

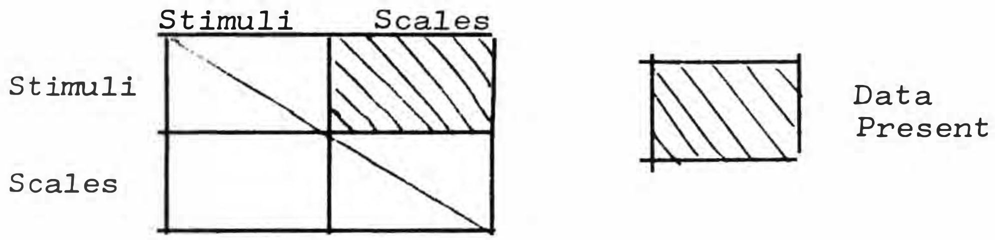
This formula appears to overcome the objection to all the previous indices that they will be too sensitive to the actual scales chosen. e.g. with a Euclidean distance formula correlated scales would have the effect of weighting some directions more than others.

Green and Rao (1972) describe a scaling of rating scale data via a linear discriminant analysis, which is equivalent to assuming a linear relation between distance and Mahalonobis D^2. This produced substantially the same configuration as the other methods, direct and indirect, on the same stimuli, the interpoint distances correlating only slightly lower than across the various Euclidean scaling methods. However, as the authors concede, this would not necessarily be the general finding. The data of this study is not particularly suitable for general comparison of direct and derived dissimilarities, since the correspondence between them is unusually high (due probably to a fortuitous choice of scales), and, in any case, the direct dissimilarities were not in fact direct, but are derived by TRICON from row conditional data.

Conjoint Scaling

Conjoint scaling refers to any procedure designed to produce a simultaneous scaling of two distinct sets of stimulus objects from data indicating some conjoint effect of an item from each set. In this context a scaling of this type can be effected by regarding rating scales as a distance between a stimulus object defined as a vector $(x_1, x_2, x_3 \ldots)$ and a notional stimulus defined by the attribute, defined as a vector of the type (0, 0, 0, ... 1, ... 0, 0). These distances can then be scaled in the normal manner by regarding the Cartesian product of stimuli x scales as coming from a supermatrix of the combined set of stimuli and scales, of which the within sets data is missing. This procedure is known as unfolding or off-diagonal scaling. The reason for the off-diagonal title can be clearly seen in Figure 1.

FIGURE 1 - OFF-DIAGONAL SCALING MATRIX



Preference scaling is another example of the off-diagonal design, where the sets are now individuals x stimuli rather than scales x stimuli. The term unfolding refers to the fact that this procedure can be considered in some sense to be a generalisation of Coomb's (1964) unfolding theory. The rating scale points and individual points in the configuration conceptually correspond to Coombs' notion of an ideal point as means of relating preferences to distances; the closer the stimulus is to the ideal point the more it is preferred.

However strictly speaking it is not unfolding as it does not use the Coombs unfolding principles to derive the solution. These are really appropriate only for unidimensional stimuli, and while attempts have been made to extend them to the multivariate case (Bennet and Hayes, 1960) the problem is really only manageable in two dimensions (see Coombs, 1979).

Rather the principle involves seeing preferences as a relation on similarities, and hence able to be mapped into relations on distances. The dimensional configuration underlying the preference judgements can thus be derived from these implied constraints on pairs of distances. This is of course a limited form of similarities data as only a small percentage of all possible pairwise distance contraints are present. However in principle a standard MDS program designed for similarities data could operate on these constraints derived from preference data, and most of the commonly used algorithms have been adapted to do this.

Coombs (1979) criticises the utility of this approach because of the very low percentage of data that is possible - in most common cases something less than 1/100th of one percent.

This implies an extremely high degree of indeterminacy, and, as Coombs points out, this indeterminacy is not equally distributed over all stimulus pair relationships.

Very few constraints are placed on the relative location of the individual ideal points, although the stimulus configuration is reasonably well determined. Thus if the research objective is to produce individual difference information then this procedure contains a high degree of uncertainty.

If the objective is to produce the dimensions underlying preferences then there is some more hope of achieving this.

Off-diagonal scaling appears to offer a major theoretical advantage over artificial distances scaling in that it is not necessary to specify the form of the function combining rating scale difference to produce distances in the reduced configuration. There are both practical and theoretical reservations relating to this technique. The procedure may not always be practical as the reduced data/stimulus ratio increases the problem of degeneracies in the weakly restricted non-metric case. The rather uncertain theoretical status of the distance measures has also to be weighed against the need to define the index of association beforehand.

The theoretical rationale for this type of scaling is clearly even more disputable than a conventional similarities scaling. A possible counter argument offsetting this however are its lower pretentions as a psychological model. Obviously if one is going to scale stimuli and people in the same space one does not expect to be representing basic perceptual processes too accurately, but rather giving a pictorial summary of the way the two sets of objects interrelate. The idea of a distance is now not an assumption of homomorphism with a psychological process e.g. similarity, but a convenient way of representing a relationship between two sets of objects It is thus a question of empirical curosity rather than theoretical necessity whether in fact a relationship exists between the configuration obtained by this method and by conventional within-groups scalings.

Young (1973) describes several off-diagonal scalings as examples of the use of his program POLYCON, and in these there is no suggestion of their being in correspondence with other configurations. Each space is interpreted only as a set of relationships across groups, not within groups. If two stimuli are located close together this is interpreted not as evidence of a direct similarity, but merely as showing they are similarly described by the rating scales.

For example, in an analyses of data on ratings of ethnic groups, the direct similarities produced three clusters, whereas the unfolding analysis produced five clusters. i.e. these clusters were sufficient to represent the perceived similarities in the group, but five were necessary to represent the way they were described by the set of adjectives.

It appears therefore that one can conclude that in general non-metric distance methods for the analysis of profile data contribute little structure that could not be obtained by normal factor analytic methods, and do not really justify their additional interpretational implications.

Rosler (1979) for example found three mode factor analysis easier to interpret in terms of known theoretical structure than INDSCAL in an analysis of three way rating scale data of German politicians.

A more promising approach to the use of profile data may be in combining it with direct similarity estimation data to produce a joint scaling consistent with both. Young (1979 a,b) described a new model denoted PRINDSCAL being developed along these lines.

CHAPTER IV

VALIDATION OF MULTIDIMENSIONAL SCALING SOLUTIONS

The central issue in the debate over the role of multivariate procedures like MDS in psychological research is of course that of obtaining a meaningful substantive interpretation of a given solution, and establishing the consistency of this interpretation with prior psychological theory. This of course raises the controversial question involved in the interpretation of all methods of multivariate analysis:- can such proceduresbe considered as psychological models i.e. representations of cognitive processes or are they simply methods of data reduction - simplified displays of patterns of interrelationships in the data. The position being taken here is that the term data reduction is potentially quite misleading - there can be no data reduction in the absence of some theoretical paradigm to guide the decision about what aspects of the data should be preserved.

This section will review various ways by which theoretical interpretations can be accorded to a MDS configuration and the means by which these interpretations can be cross validated with other applications of the substantive theory.

Dimension Based Interpretations

By far the most dominant means of interpretation has been to regard the dimensions of the space as basic variables and attempt to match them with substantive theoretical constructs. A psychological modelling approach would require that we can interpret both the dimensions and the combination rule used by the procedure as a direct analogy of the cognitive process which generated the data. A less stringent (and more realistic) aim would be to interpret the dimensions as primitives of a basic perceptual process on which more complex cognitive processes are predicated. If the recovery of the precise nature of this process is the object of the research then more direct model testing procedures than MDS would of course be preferred. In addition, from the arguments considered in the previous chapters, it seems reasonable to regard the case against MDS as an appropriate model for either similarity of preference judgements as being fairly convincingly demonstrated, if not yet completely accepted. However, this does not preclude the second more restrictive aim of using MDS to recover the structural aspects that underlie these processes.

The cognitive operations on this structure would thus represent unknown functions which are to be accommodated by the indeterminancy of the scaling procedure. While strictly speaking this implies certain restrictions such as independence of dimensions and monotonicity between the distance function and the true combination function there is evidence that most procedures are reasonably robust to substantial departures from these requirements. This approach, while less than complete modelling, is nonetheless more than just data reduction, as some attempt can still be made to provide a meaningful substantive interpretation that has some theoretical relevance i.e. that could be predicted on an a priori basis, rather than simply a post hoc description of structural aspects of the given data.

Attempts to demonstrate that the dimensional orderings extracted by MDS procedures can be provided with at least this limited interpretation of psychological validity have generally followed one of three main types of procedures. A distinction will first be introduced between internal validation, which compares the solution with known aspects of the input data, and external validation, which utilises information additional to that used to produce the configuration being validated.

Internal validation procedures have followed two main lines:

1. the data are generated by subjects using stimuli of known (and usually limited) dimensional characteristics, for example abstract patterns constructed to vary only in a certain number of aspects.
2. the data are generated artifically from a known configuration with a specified combination rule and specified degree of error.

Both these procedures enable the performance of the MDS procedures to be evaluated under conditions where the expected configuration is known in advance. The Monte Carlo studies of course provide the greatest knowledge of the actual conditions governing the generation of the data, and these can be specified in terms of degree of error on a specified measurement model. The robustness of the scaling procedure to both appropriate and inappropriate measurement models can thus be evaluated. With the subject generated data the cognitive strategy is unknown, but the possible dimensions on which it can be predicated are known and restricted to manageable proportions.

This provides empirical test data with still some degree of control over its expected form. It is of necessity however restricted to stimulus sets with quite discrete, perceptually easily discernible, dimensions, a characteristic previously shown to be of some significance and thus would not necessarily be comparable to data generated from more realistic stimuli.

External scaling places the fewest restrictions on the possible data types one could use for the initial scaling provided there is some theoretical or empirical information one could use to predict at least some aspect of the resultant configuration. That is, it aims to test the geometric representation assumption by construction of at least some components of the structure from separate information. These may range from single dimensions, individual dimension weights to the total configuration.

External Validation of Dimensional Interpretations

Probably the most common method of validation of the dimensions of configurations obtained by MDS methods has been by comparison with the known physical attributes of the stimuli. Even in cases where the configuration is not known an intuitively reasonable result in terms of the physical dimensions is often regarded as proof that the scaling program has performed correctly. We will not devote space here to reviewing all MDS which have shown such a correspondence between the solution and the known characteristics of the stimuli; suffice to say that plenty do exist. The study most often quoted is the tea tasting experiment of Carroll (1972). Some of the problems involved in this type of validation were detailed in the previous chapter.

The situation to be considered here is the cases where the physical dimensions of the stimuli are either not known or where there are a very large number of possible attributes which could be used as the basis of a perceptual strategy. That is, if one uses real world rather than abstract stimulus definitions then an exact identification of the attributes which act as basic determinants of the subjective strategies becomes the dominant theoretical problem.

A reasonable number of studies have reported significant success in matching co-ordinate values on dimensions from MDS of real world stimuli with corresponding values on some externally derived scale.

Burton (1972) used similarity judgements on 60 occupation names and extracted 3 dimensions, which appeared interpretable as independence, prestige and skill. He then asked subjects for paired comparisons judgements on 'prestige', 'income' and 'social status'. These judgements were found to have a very good fit with the prestige dimensions. 'Prestige' and 'status' were highly intercorrelated (.910) and correlated .904 and .917 with the MDS prestige dimension. Thus at least one dimension of this solution was found to be externally definable and measurable.

Wish and his colleagues (Wish 1970, 1971, Wish Deutsch and Biener 1970, Wish, Kaplan and Deutsch, 1973) describe a series of comparisons across similarity preference and semantic differential data on the perception of nations. The similarities were incomplete and were first scaled by MDSCAL, and the distances from this were input to INDSCAL - a procedure we will charitably describe as 'novel'. 4 Dimensions were extracted, and these were all correlated with the mean ratings on the rating scales. Multiple correlations were also calculated with each scale, and most were fairly high (.7 to .9) indicating that most of the information from a large number of rating scales was contained in the 4 dimensions. There was less agreement between similarity and preference spaces although one would not expect those to be identical.

Cliff and his co-workers (Cliff & Young, 1968 a,b) found a substantial degree of external validation for a variety of stimuli. Photographs of facial intensity were rated on similarity and intensity of emotion. (Cliff and Young 1968). A TORSCA scaling of the similarity judgements produced a V shaped solution in 2D, conforming to that obtained in previous studies (e.g. Shepard 1962). This is usually interpreted as intensity versus affect. The intensity ratings as expected correlated highly with the intensity axis. Cliff and Young also tested the hypothesis that the intensity ratings correlate with distances from an origin in space. Such an origin was found with a rank order correlation of .97 between intensity and distance. It was not made clear just why ideal point type representation was preferred, especially since it was located at one extreme of the stimulus configuration, making the rank order of the distances from it virtually identical with those of the projections on the intensity axis.

Another study used trait description adjectives, which were rated on similarity and favourableness as descriptions of two types of people; a captain of a ship and a technician. The dimensions of the 3D solution were used as predictors in a multiple regression with the favourableness ratings. The multiple correlation for both was .96. A third study used simulated air raids as stimuli. These were rated on similarity, degree of threat and best course of action. The multiple correlation between the 4D solution and threat ratings was .98, though this was almost entirely due to one dimension. A discriminant analysis indicated a high degree of predictability of modal courses of action from the configuration. Much less success however was met in attempts to correlate items scores on personality questionnaires with configurations derived from ratings of perceived meaning (Cliff, 1968 a).

Cliff and Young claimed that these results provided definite proof of an internal representation corresponding to the MDS configuration. They argue that 'an individual's responses to members of a set of stimuli are based on his internalized conception of them, and that this internalized conception can be revealed by multidimensional scaling analysis of interstimulus similarity". (1968, p.269).

While one can hardly concur with the sweeping conclusions of Cliff and Young it is clear that dimensions from MDS configurations can be related to independently derived measures of theoretical contructs and this contributes to the validity of the solutions as reasonable models of the perception of those stimuli.

Another type of validation study has attempted to demonstrate a high degree of consistency in the configuration derived across variations in the nature of the cognitive judgement required in order to provide evidence for a constant cognitive structure.

For example, Henley (1969) collected five types of information on names of animals to obtain evidence for a constant semantic structure. The first was a conventional paired comparisons similarity task. This was first analysed by an unspecified points-of-view analysis to obtain a homogenous group of subjects. This eliminated 3 out of 21 subjects. The rest were averaged and scaled by TORSCA, producing 3 dimensions; size, ferocity and humanness. Four other types of information collected were then related to this. These were free recall, triad similarities, associations and paired associate learning.

In the free recall task subjects were asked to write down all the animals they could think of. A matrix of average intra-list distances was calculated to test the theory that this would be related to psychological distance. While very similar pairs did tend to be written down together the entire matrix correlated only 0.17 (ns) with the TORSCA scaling. This is however only a very indirect measure and this low correlation is not surprising.

Two experiments used Torgersons (1958) method of triadic similarities for subsets of 6 and 12 animals. The correlation between the pair and triad rating matrices was .90 for the 6 animals subset, .93 for the 12 animal subset. There were some differences in the configuration resulting from TORSCA scalings of these e.g. the 6 animal subset produced a solution in two dimensions, only one of which was recognizable as an original dimension. This is of course to be expected with lower numbers of stimuli and hence lesser stimulus variation.

Henley also collected information on association and paired associate learning tasks, to see if association and generalization were related to similarity structure. She found evidence for association but not for generalization, the latter being interpreted as due to interference from the serial position effect.

These comparisons are obviously of two different types. The triad-pair comparison was one of method only, while the listing, association and generalization experiments were comparisons of structure across cognitive processes. The results Henley obtained were not precise enough to confirm or reject the notion of a constant semantic structure.

Many authors have examined the possibility of a direct relationship between multidimensional structure and the subjective lexicon.

Miller (1969) argues that while there is a relationship to some degree, this can be at best an approximation. He asked subjects to form a free sorting of 48 English nouns and found that the incidence of sorting together obeyed the requirements of a metric.

Miller distinguished three main types of semantic organisation. When sorting two items together some features must be ignored. Only distinctive features are relevant, as only these can be ignored by different numbers of judges. When all the items in the set have values for every feature then we have a paradigmatic system e.g. kinship. When they form a sequence then one has a linear organisation. The final type is a hierachical organisation. Miller showed that a multidimensional interpretation was appropriate for a paradigmatic or linear organization, but not for a hierachical organization.

Fillenbaum and Rapoport (1971) also showed the inadequacy of MDS for dealing with semantic structures incorporating hierachical organization. They compared MDS and Johnsons (1967) hierachical clustering procedure for a number of sets of semantic structures; e.g. a list of colour names, and a list of verbs related to the HAVE family.

Proximity data was obtained by three methods - tree construction, complete undirected graphs and direct grouping. While a MDS solution quite adequately accounted for the data from the colour word set, and indeed seemed superior to the clustering analysis, it was shown to be clearly inadequate for the HAVE word set. While the 2D solution approximately reproduced the same clusterings no interpretation could be given to the dimensions, nor to the way the groups were distributed in space. On the other hand the cluster analysis did yield intuitively reasonable groupings, and the hierachical organisation made it possible to say what governed assignment to a cluster and what differentiated between clusters.

A more direct method of examining the relationship between similarity and semantic structure is provided by Rumelhart and Abrahamson (1973). They tried to find a cognitive process which would have a direct analogy as a transformation in a multidimensional semantic space. Analogical reasoning itself was considered to be such an analogy. If an analogy is defined as an ordered triple (A:B:C; A is to B as C is to ?) then the process of finding the analogy would have a logical counterpart in multidimensional space of a projection from C parallel to and of the same length as AB.

Rumelhart and Abrahamson used Henley's data to define a space of animal names in which to test this assumption. They found a very good fit to the hypothesis that the closest stimulus to the point predicted by the parallogram transformation is most likely to be picked to complete the analogy, and that probability of choice decreases with distance from this point. To predict this a bit more exactly they combined Luce's choice axiom and Shepard's theory of generalization and derived an exponential function of the distance relative to the distance to all alternatives as the estimate of the probability. This produced a correlation of .937 between the predicted and observed number of subjects choosing each alternative. Luce's axiom was naturally extended to obtain the probability for a second choice etc., again producing quite a reasonable fit.

Thus to summarise, a large number of studies have been able to obtain significant correlations between MDS configuration dimensions and information derived from independent considerations relating to the substantive psychological theory. However this is clearly far from proof that any real correspondence can be assumed to exist between such configurations and any general conception of cognitive structure. The concept of goodness of fit by itself is not a very adequate criterion with which to resolve such theoretical issues. Correlation coefficients between psychological measures must always be limited by the far from perfect reliability of the data. Thus it is highly unlikely that one would be able to tell by goodness of fit alone whether a given representation of data is a valid psychological model or just a good approximation. One has no basis to determine whether a given departure from fit represents random noise or a real substantial failure of the model. Clearly then one needs a more falsifiable method of theory evaluation by which one can show that a given theory is not only a good approximation but better than any alternative hypothesis.

An additional limitation with these dimension-wise methods of validation is that they do not test whether the solution is sufficiently comprehensive to cover the domain of the substantive theory. It is in fact a frequent conclusion from results of the application of MDS to a given domain that the solution, while often showing some correspondence with known substantive theory, is typically limited to the most obvious and broadly defined dimensions. Fillenbaum made this observation (personal communication) after an exhaustive series of studies of the use of MDS to explore semantic structure

(Fillenbaum and Rappoport 1971). The extreme contrast between the level of detail in corresponding linguistic theory, and the very crude approximations produced by the MDS configurations could hardly be conclusive to any attempt to interpret such configurations as representations of cognitive structure.

Restricted Multidimensional Scaling

One approach to the use of data analysis in a theory evaluating context which has recently become popular is the concept of confirmatory analysis. The distinction between exploratory and confirmatory data analysis (Tukey 1962, Kaiser 1970, Hubert and Subkoviak 1979) relates to the extent specific theoretical hypotheses are considered prior to application of the analytic procedure. An exploratory strategy involves the use of an analysis technique on a given data set with the aim of identifying interesting relationships, patterns, and the like. Alternatively, a confirmatory approach involves the testing of an a priori hypothesis that is generated from a source distinct from the data to be used for the purposes of validation.

The term confirmatory refers to a more limited type of hypothesis testing than traditional experimental paradigms, as the degree of support for the hypothesis can only be expressed in correlational terms.

Confirmatory data analysis procedures have been most extensively developed within the framework of factor analysis (Joreskog 1969, 1970, 1971; McDonald 1976, 1978) and normally involve the a priori specification of some of the parameters in the factor model. This confirmatory analysis principle has also been developed recently within the framework of multidimensional scaling, with procedures being described by Carroll, Green and Carmone (1976), Noma and Johnson(1977), Bentler and Weeks (1978), Bloxam (1978), Borg and Lingoes (1978a, 1979, 1980). These procedures essentially involve estimating the stimulus configuration while at the same time constraining the model in a way appropriate to the stimulus design. If the model with constraints fits nearly as well as the model without constraints we can accept the restricted model and attribute differences in the unconstrained model to random error.

Bentler and Weeks (1978) describe a class of restricted multidimensional scaling models in which certain of the parameters can be fixed as constants, a range of values, or as proportions of other parameters. Their developments have been initially limited to the standard orthogonal Euclidean model, and thus the parameters are the dimension co-ordinates. In the completely unrestricted model the number of parameters is dn where n is the number of stimuli, and d the number of dimensions (although slightly less than this number are free to vary, for example one parameter per dimension is needed to fix the centroid). Thus parameters may be added only in multiples of n, and only a very crude trade off between fit and number of parameters may be obtained. One is, for example, often forced to keep to low dimensional solutions in order that reasonable ratios of stimuli to dimensions may be preserved. However if accurate prior information is available that enables one to apply a restricted model with fewer free parameters higher dimensional models can be considered without such a corresponding loss in determination. The fit of the model can be meaningfully compared across alternative specification with equivalent numbers of free parameters, independent of dimensionality.

Bentler and Weeks quote an example of fitting Ekman's (1954) colour data in which a third dimension was added, but constrained to be proportional to physical wavelength. This therefore required only a single additional parameter. This model reproduced the colour circle in the two unrestricted dimensions, but the third dimension allowed a separation between the ends corresponding to psychologically similar, but physically distinct red and purple tones (i.e. it produced a spiral, rather than a circle).

Bloxam's and Carroll, Green and Carmone's models both extend the confirmatory procedure to fitting individual, rather than common, stimulus configurations. Bloxam's model is the most general, as it allows for non-orthogonal axes, as well as individual differences in the configuration. The model provides for simultaneous scaling of n stimuli in N spaces under specified constraints of correspondence between the N spaces. The N spaces can of course relate to differences in experimental conditions, as well as interindividual differences.

Bloxam's model is based on the most general form of the Euclidean distance, which incorporates individual weightings along dimensions not necessarily restricted to be orthogonal. That is, given an m dimensional configuration of p stimuli, the generalized distance between stimuli j and k in individual space i is denoted by

$$d^2_{jk.i} = (\alpha_{ji} - \alpha_{ki})' \; C_i(\alpha_{ji} - \alpha_{ki}) \qquad (18)$$

where α_{ji} is the j^{th} coordinate vector of projections on m' ($\leq m$) reference axes in space i and C_i is an $m' \times m'$ symmetric positive semidefinite (rank m) matrix for space i.

An appropriately normed C_i contains off-diagonal entries corresponding to the direction cosines of the angles between the m' reference axes.

Equation 18 is thus a slightly different application of Equation 17. It represents a generalization of Equation 10 incorporating a nondiagonal C_i as in Equations 7 and 13.

Constraints on this model can be expressed in terms of N matrix equations of the form

$$A_i U_i = X_i \qquad (19)$$

where A_i is the matrix of projections of p stimuli on m' reference axes, U_i is the $m' \times m'$ matrix containing the non-null columns of the triangle Gram factor of C_i, and X_i is the $p \times m$ matrix of projections on m orthogonal reference axes. The elements of A_i and U_i can be fixed constrained by certain functions, or free to vary.

Carroll, Green and Carmone's (1976) model involves a slightly less restricted class of constraints in which the projections on each of m orthogonal axes are constrained to be a linear combination of a priori measures of the stimuli.

Carroll and Pruzansky (1977) extended this principle to the individual differences case where one has a linear combination of idiosyncratic measures for each subject. This can still be considered as a restricted case of Bloxam's model.

This confirmatory testing principle appears to offer significant advantages over other methods of validation e.g. the more traditional approach of trying to locate a reference axis corresponding to each factor of the design such that stimuli representing the same level of the factor have the same projection on that axis.

This latter approach has two significant disadvantages. Firstly, reference axis can seldom be located meeting these exact conditions, and secondly, even if an axis can be located approximately meeting the conditions it is not necessarily the case that the axis provides a simple measure of the factor's effect on the configuration. One cannot be sure whether the differences are due to random error or some non-random variation in the factor's effect (e.g. its interaction with another factor) on the configuration.

Limitations of Dimensional Interpretations

Several theorists have argued strongly against the undue emphasis typically placed on dimensional interpretations. Lingoes (1977 a, 1979) for example feels that a dimensional interpretation is rarely justified and argues that external continua should instead be mapped into the more general concept of directions.

The distance model simply identifies points in an arbitrary reference frame and there is therefore no logical necessity for the dimensions of any particular solution to correspond to theoretical orderings. Since the configuration is only determined up to an arbitrary rotation it is not suprising that conflicting interpretations often arise, as a different rotation of the axis will emphasize somewhat different aspects of the solution. It is thus rather hard to reconcile attempts to interpret dimensions as a reflection of some underlying reality with this lack of rotational unique ness. There is no logical reason why any set of realistic or useful dimensions must be limited to being orthogonal and to the minimum number required to spatially represent the interpoint distances.

Thus it is clear that MDS cannot be considered as a simple extension of unidimensional measurement as there is this additional element of subjectivity corresponding to the lack of any firm empirically based rules for determining which set of numbers are to be assigned to which dimension.

Identifying Directions For Dimensional Interpretation

The technique of inserting externally measured scales as explanatory vectors in a MDS configuration has long been a common interpretational device in standard dimensional analysis. These vectors can be expressed as a weighted combination of any given set of orthogonal dimensions, the weights corresponding to the projections on the co-ordinate axis, and their location is therefore independent of the particular rotation used to specify the configuration. They can be obtained by multiple regression procedures using the dimensions as predictors.

These vectors may either be used as objects of interpretation in their own right, or more commonly, as aids in the interpretation of dimensions according to their correlations with the total set of explanatory vectors. Additional variations which assist this interpretation are to use nonmetric regression, and to plot vector lengths according to the size of the multiple R with the total set of dimensions. Several computer programs are available for using this procedure (Carroll and Chang, 1964; Lingoes, 1968 a,b), and a number of investigators have supported its use as an aid to dimensional interpretation (Green and Rao 1972; Rosenberg and Sedlak, 1972). Lingoes (1977 a) however, observes that these methods should not be applied from the point of view that each dimension must have an interpretation, and questions the tendency to assume equivalence between a dimension and a closely aligned vector when the variance in common is not sufficiently high.

Non-Dimensional Interpretations

The difficulties associated with the interpretation of spatial dimensions suggest a need for an approach to interpretation of MDS configurations which relies on information that is invariant under rotation. As Lingoes (1979) observes, by using formal aspects relating to the pattern of configuration of points in the space, rather than the dimensions per se, as they reveal orderliness or some law of formation in this pattern one can to some degree side-step the issue of nonuniqueness. The use of directions as described above can contribute to this independent interpretation to some extent. However a more promising approach appears to be in terms of direct classification of the geometric form of the pattern of points. Definitional systems for such patterns have been developed based on the concepts of clusters, regions and manifolds.

Regions and Clusters

The most elemental nondimensional approach to the interpretation of spatial structure is by partitioning the space into regions such that all points within a region possess similar values on some specified set of theoretical features. A region can be defined most generally as a set of points satisfying certain contiguity relations (Lingoes 1977 b). Most partitioning procedures have used a slightly more restrictive definition in terms of clusters, based on some criterion identifying strong relationships within sets and weaker relationships between sets. The use of clustering procedures has long been a fairly standard aid to the interpretation of MDS solutions, with the most popular procedure being the hierarchical clustering program of Johnson (1967). This uses a very simplified nearest neighbour clustering principle appled at a series of levels relating to progressively varying the cutoff point demarcating within sets and between sets distances. While this has been supported as an aid to theoretical interpretations in recent reviews of MDS procedures (Shepard 1974; Carroll, 1976) it is still clearly heavily dependent on subjective assessment of the identification of the theoretical relevance of the series of clusters. There is little scope for formal rules for prior determination of what are theoretically distinct clusters, and with such a variety of possible clusterings both within and between procedures, there seems insufficient structural information of sufficient stability to be of any significant theoretical value.

Manifolds and Facet Theory Interpretations

For non-dimensional spatial representations of empirical data to be usefully related to substantive theoretical interpretations they clearly need to meet two important conditions:

(i) formal aspects of the empirical structure must be sufficiently well defined to enable stable empirical regularities to be unambiguously identified.

(ii) these formal descriptions must be capable of being related to corresponding aspects of some theoretical definitional system.

That is, one must be able to establish a parallel between empirical and theoretical definitional systems, with some clearly defined rules for a mapping between them. For this we require what Foa (1962 p.166) describes as a metatheory that provides rules for making the transition from conceptual scheme to empirical structure. This implies a need for the substantive theory to be specified in terms which are compatible with the intended representation of empirical results. One such theory which has been developed as a counterpart to MDS configurations is Guttman's facet theory (1957). Guttman in fact expresses the need for all theory to meet this metatheoretic requirement of including both theoretical and empirical structures. He defines a theory as "a hypothesis of a correspondence between a definitional system for a universe of observations and an aspect of the empirical structure of those observations, together with a rationale for such a hypothesis". (Gratch, 1973 p.35).

Facet theory aims to provide a unified scheme for defining a universe of observational items and deriving hypotheses relating to their predicted spatial representation.

The theory uses a technique known as mapping sentences to provide a general means for constructing and presenting definitional systems for universes of observations. Mapping sentences constitute theoretical statements containing a number of facets, each of which can take on a number of qualitative values. That is facets can be regarded as the set of qualitative variables required to completely describe a given empirical system. Any given observation can thus be described as one element of the Cartesian product of the facets (denoted a structuple). Mapping sentences are thus used to express hypotheses relating to the prediction of behaviour throughout an entire conceptual domain. For example a simple mapping sentence relating to the operant conditioning principle is listed below:

$$\text{A } \overset{A}{\begin{pmatrix}\text{high}\\ \text{low}\end{pmatrix}} \text{ average rate of reinforcement that}$$

$$\text{is } \overset{B}{\begin{pmatrix}\text{constant}\\ \text{variable}\end{pmatrix}} \text{ in terms of } \overset{C}{\begin{pmatrix}\text{number of reinforcements}\\ \text{intervals between reinforce-}\\ \text{ments}\end{pmatrix}}$$

$$\text{under } \overset{D}{\begin{pmatrix}\text{high}\\ \text{low}\end{pmatrix}} \text{ deprivation produces } \overset{E}{\begin{pmatrix}\text{high}\\ \text{low}\end{pmatrix}}$$

rate of the response reinforced in facet A.

The S - R paradigm has thus been expanded using facets to specify all relevant variations in the experimental conditions (they may relate to subjects, stimuli, conditions, etc.) which are theoretically related to an ordering on the predicted strength of response. The corresponding ordering on the structuples induced by this responses ordering constitutes the conceptual structure of this particular empirical domain. Guttman derived a series of principles providing general rules relating the conceptual structure obtained by facet description to both the structure of the similarity matrix and the resultant MDS configuration. That is one requires metatheoretic principles enabling one to move from logical or conceptual structure to an empirical structure.

The most obvious and basic of these is the contiguity principle (Guttman 1959). Expressed in its most general terms (Foa 1958) this states that items that are closely related in logical space should be closely related in empirical space. Thus stimuli which are similar in their facet structure should be closely related in geometric space.

One thus requires a definition system of empirical structure which is expressed in terms of the contiguity of subuniverses of items. Guttman developed a description of structure in terms of ordered regions known as manifolds. He derived precise definitions for two basic types of manifolds denoted simplexes which represent simple orderings of subuniverses of items, and circumplexes, which represent recursive (i.e. circular orderings). More complex structures can also be specified as combinations of these two basic types for example, a duplex is a combination of two linear orderings, a radex, a circular and a linear ordering.

Facet analysis thus represents a system for deriving empirical hypotheses concerning partitions of the obtained space into contiguous regions each containing variables of a class conceived in advance by the definitional system. That is, the prior specification of the role played by a given facet maps into a regional hypothesis in the MDS space.

There are however certain limitations on the degree to which this can be done completely on a priori basis. For example the derived facet structure is not independent of the ordering of facets. A predicted circular order for a set of three facets corresponds to a circumplex, with two of the facets as semicircles within the circumplex with the third in four segments. Altering the order of the facets means a different one will play the more chopped up role, and it thus becomes a psychological problem to determine which two of the three should play the dominant roles in the representation. Foa (1962, 1965) used the term semantic principle components of the proposed order to treat this problem. More detailed discussion of the representation of facet structure are given in Borg (1977), Shye (1978 a,b,) Shepard (1978).

Some examples of the successful application of facet theory to the interpretation of MDS solutions are given in Lingoes (1977b), Shepard (1978), Just and Murray (1978). For example Shepard describes several circumplex interpretations for, as well as the familiar colour circles, musical tones and the well known speed of mental rotation data (Shepard and Metzler 1971).

An important methodological problem related to this approach is the need to be able to identify the partitions to form the appropriate manifold. Shye (1978 a) refers to the desirability of doing this in a more quantititative fashion, employing a loss function involving not only technical goodness of fit but also deviations from a substantive regional hypothesis. Some progress in this direction is reported in Lingoes and Borg (1978 a) with the development of CMDA (Constrained/ or Confirmatory Monotone Distrance Analysis) which enables the testing of specific regional hypothesis. This program now includes a recently developed statistical test for the significance of constraining the solution to conform to the hypothesised manifold, or any other basis for regionalisation (Lingoes and Borg 1980). A closely related procedure to this was proposed by Guttman under the title of faceted SSA (Smallest Space Analysis) (Guttman Note 2). A necessary condition for these tests is a formal definition of the concept of contiguity. A variety of ways of doing this are discussed by Lingoes (1979). Lingoes proposes that rather than decide on a single form one should instead utilise a series of definitions of increasing strength. This provides a basis for a somewhat flexible regional analogue of the confirmatory multidimensional scaling procedures developed for testing dimensional hypotheses.

By using varying definitions of contiguity of increasing strength one can assess the degree of disjointness of the empirical partitioning implied by a particular substantive theory. An example of the application of this technique is given in Ronen, Kraut, Lingoes and Aranya (1979). This analysis relates to the extensively studied problem of deriving the most appropriate taxonomy of work motivation. Previously, mainly factor analytic studies had produced conflicting results with most support for Herzberg's two factor theory (Herzberg et al 1959), with little support for Maslow's widely accepted fivefold hierachy of motivation. Ronen et al however showed that Maslow's theory can be perfectly fitted to a MDS configuration of some highly reliable data of attitudes to work in terms of the partitioning produced by a series of parallel lines.

All three major theories (Maslow's, Herzberg's and Alderfer's (1969) three component theory) were consistent with the data in terms of corresponding to identifiable regions. However the comparative analysis referred to above showed Maslow's to be substantially more powerful in terms of the number of contiguity constraints it implies which were satisfied in the configuration.

Conclusion

To summarise then, one must conclude that the interpretations produced by most applications of MDS procedures to a given empirical domain have little real value to contribute to the substantive theory in that area. While a crude correspondence can usually be established between empirical and theoretical aspects this is typically on a very basic level. A MDS configuration clearly can provide neither the accuracy of detail nor the comprehensive coverage required for it to be able to be used as a means of deriving empirical structures which can add significant new information of any theoretical value. There does seem more promise however in its use in a more limited theory testing role, to confirm the extent spatial constraints implied by previously derived substantive theories can be satisfied in a given configuration. As Skarabis (Note 3) points out most applications of MDS (and other multivariate techniques) are typically 'closed' in the sense that the investigator cannot guide the analysis in the light of knowledge or hypothesis. While the procedure may seek the 'best' solution it is quite likely that other solutions with quite distinct theoretical interpretations may be equally good.

One must therefore provide for a combination of both statistical and theoretical criteria to guide the derivation of geometric representations.

PART THREE

EMPIRICAL EVALUATIONS

The following series of investigations were designed to consider the applicability of MDS procedures to some examples of empirical proximities data, the possible interpretations that could be placed on the spatial configurations derived, and the means by which the theoretical adequacy of these interpretations (as theor etical models) could be evaluated.

The commonly used paired comparisons magnitude estimation of similarity was used, with the size of stimulus sets restricted to limits which enabled complete presentations. This was to avoid any of the theory-weakening designs involving missing or pooled data, and retain the option of performing metric or non-metric analyses.

The stimulus sets selected for these studies were designed to encompass a diverse range in terms of the types of perceptu al strategies that might be applied to them. As commented previously while simplified abstract designs with clearly distinct and dimensional components make the validation process much easier, there is a danger that these will be unrepresentative of real world perceptual processes.

The stimuli used in Experiment I were abstract stimuli, but of a class that previous research has suggested present some difficulties for the MDS models. These are circular or directional stimuli, i.e. those involving an angle or direction. The present stimuli involved a constant triangular design which varied in its angle of orientation.

Experiment 2 in contrast involved a stimulus set selected from a real world domain; emotion labels. As well as representing the more realistic case where the basic perceptual dimensions are neither obvious nor distinct it also represents an area where there is as yet no commonly accepted theory to predict what these should be.

CHAPTER FIVE

THE PERCEPTION OF SPATIAL ORIENTATION

This investigation was based on a stimulus set used in an unpublished study by Gregson (Note 1). This involved a series of isoceles triangles bisected along the vertical axis, one side of which was filled with a black brick pattern. The triangles varied only in their angle of rotation i.e. the orientation of the centre line, and in which side of the triangle was filled in. Figure 2 shows an example of a typical stimulus pair.

Stimuli were constructed in steps of 18° rotation, giving a possible total of 20 rotation angles, each with two black/white reversals. Gregson's study involved a selection of 24 stimuli from this total set of 40 possible combinations.

Gregson characterised the stimulus series as representing a closed group under the two transformations (18° rotation and b/w reversal) used to generate the series. He attempted to relate the similarities on a pair of stimuli to the combination of transformations required to turn one into the other.

Using a modified version of Gregson's notation we will denote g_1^r to represent rotation through 18° r times, and g_2^1 to represent black/white reversal.

The zero transformations are thus defined as g_1^o, g_2^o respectively. It can thus easily be seen that the two sets of transformations both constitute a group under a combination relation (i.e. performing them sequentially). The periodicity of the group is defined as the number of transformations required to convert a stimulus back to itself. Thus g_1 has periodicity 20 (or 360°) and g_2 has periodicity 2. The half-periodicity thus represents the maximum number of transformations

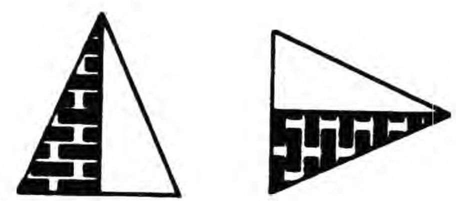


FIGURE 2: A typical stimulus pair (from Gregson, Note 1)

required to convert one stimulus into another (including the inverse transformations).

Gregson hypothesized that the cognitive processes used by subjects to evaluate the similarities of any stimuli generated by closed group transformations, such as the circular and alternating transformations used here, could be represented by a mapping onto the number of these transformations. However, as he points out, their subjective strategies need not be confined to them in a simple manner. For example he suggests a third transformation g_3, a rotation of the triangle 180° about the baseline (i.e. a flipover operation) which may be used in preference to the equivalent combination of g_1^{10} and g_2^1. This would thus represent a substantial increase in similarity at rotation differences of 180°.

Gregson attempted to accommodate individual differences in subjective strategies by alterations in the periodicity of the group. For example halving the periodicity of the group of g_1 is achieved by redefining m, the number of transformations as m' where

$$m'(g_1^k) = \min[m(g_1^k), m(g_1^{n/2-k})]$$

This corresponds to defining stimuli related by $g_1^{n/2}$ as equivalent, and thus for a given stimulus the number of transformations into the nearest of these is the determinant of similarity.

Gregson did not develop the rationale for such alterations in periodicity, although he does observe that the use of g_3 may be approximated by a better fit with periodicity 180°.

There does not appear to be any simple explanation for this in the general case of closed group transformations. Plausible explanations can however be devised in the specific case of spatial orientation. In this case transformations are rotations of a given amount in one direction, with the inverse transformations rotation in the opposite direction.

Thus if the similarity function over angular difference is defined as $f(\theta)$ we assume

$$\Big[f(\theta) \Big]_{0}^{180} = \Big[f(\theta) \Big]_{360}^{180}$$

This distinction would in fact only be defined for rotation in a constant direction for given start/target identification of the stimuli (most plausibly based on a fixed left/right relative positioning). While these effects may be plausible they were not considered here, and would be balanced out by the variable left/right positioning of stimuli. We thus define θ to represent the minimum angular difference in either direction, and it will fall in the range $0^{\circ} \leq \theta \leq 180^{\circ}$. We will assume f is some monotonically decreasing function within these limits.

An alteration in the periodicity is equivalent to a folding of the angular difference scale at its midpoint. A periodicity of 180° thus corresponds to folding the scale at 90° and equating angles increasing from 0° to 90° with angles reducing from 180° to 90°. That is we now have some monotonic function f', such that

$$\Big[f'(\theta) \Big]_{0}^{90} = \Big[f'(\theta) \Big]_{180}^{90}$$

Such an identity is clearly produced by defining θ' as the *minimum* angle between the undirected centre lines (i.e. regardless of direction). A function monotonically decreasing over θ' will thus decrease over the range $0^\circ \leq \theta \leq 180^\circ$. If one assumes the shapes of $f(\theta)$ and $f'(\theta)$ to be equivalent then f should be a constant proportion of f' rescaled by a factor of two, i.e.

$$f'(\theta) = \alpha f(2\theta)$$

Equivalent slopes would imply

$$f'(\theta) = \tfrac{1}{2} f(2\theta)$$

Similarly a periodicity of 90° equates with folding this scale again at 45°. This appears to represent a strategy based on the angular difference between the stimuli or the difference between the stimuli or the difference between this and some number of right angles; a strategy with only a limited degree of plausibility if one stimulus is in a familiar orientation regarding such regularities (e.g. vertical).

Gregson did find some subjects showed a better fit for periodicities of less than 360° but did not offer any interpretation of this. A comparison of the slopes and intercepts of the regression lines (at the periodicity of best fit) indicated a significant difference in the intercept but not the slope (using only subjects with significant regressions). This implies that g_2^1 contributes an independent additive effect to dissimilarity, while g_1^k contributes an effect which is a constant ratio of K, defined at a periodicity which may vary over subjects and the value of g_2.

The Experiment

This investigation is a more extensively analysed report of a previously published study in which only the first set of results were included. (Fraser, 1976). This involved a replication of Gregson's method in conjunction with the application of a MDS procedure. This study attempted to compare the two alternative explanations for the increased similarity at 180°. The 2θ strategy considered above in fact corresponds more directly with what is being tested by the regression with periodicity 180° and thus cannot be distinguished from Gregson's g_3 by this method. It was hoped however to distinguish these strategies by an interpretation of the MDS configuration derived from the similarities.

Procedure

Seventeen subjects made similarity judgements on all pairs of a set of 12 stimuli, presented in randomised order, balanced for order within pairs. Subjects were first year undergraduate psychology students and were tested individually. Stimulus pairs were projected onto a 1m x 1m screen for a constant duration of about 8 seconds, and subjects gave their rating verbally on a scale of 1-100. Examples of different and identical stimulus pairs were presented during the instructions; subjects were told to score identical stimuli as 100, but were not told any number to pair with any non-identical pair.

Instructions to subjects

The following instructions were read to subjects prior to the experiment: "You will be shown pairs of patterns side by side on the screen, and you are to give a number between

0 and 100 to each pair of patterns. This number is to represent a point on a scale expressing how similar the two patterns are as you see them. "100" means they are the same, "0" means they are not at all similar. Any number between 0 and 100 is possible. You call out your answers as soon as you can.

"I will show you some examples of pairs of patterns and the numbers you should give them:- These are the same and therefore should be scored 100 - these are not the same and therefore should be scored some number less than 100. Apart from that it is over to you to decide how to allocate the numbers ....."

The subjects then received some other instructions regarding reaction time and confidence measurements which will not be considered here.

The Stimulus Series

The stimuli used in this study consisted of a set of 12 isoceles triangles of the type used in the Gregson (Note 1) study. The set of 12 stimuli used were selected from the total of 40 possible as detailed in Table 2.

Results

The data from all 17 subjects were simulta neously analysed using the metric weighted Euclidian model individual differences scaling program INDSCAL (Carroll & Chang 1970). Some shortcomings of this model were discussed previously, although most of these considerations were unavailable at the time of this study.

TABLE 2

Specifications of Stimuli: Experiment One

No.	Black/White Code	Rotation from Vertical	
		Degrees	Units
1	0	324	18
2	0	0	0
3	1	18	1
4	0	72	4
5	1	90	5
6	1	126	7
7	0	144	8
8	0	180	10
9	1	198	11
10	0	252	14
11	1	270	15
12	1	306	17

The configuration derived from INDSCAL produced a reasonable fit and interpretable solution in 5 dimensions. Graphs of some of the 2-D subspaces of this 'common space' configuration are shown in Figures 3-10 . Dimensions 1 and 2 clearly correspond to θ, the angular difference, and Dimension 4 separates the stimuli into two groups corresponding to the black/white reversal equivalence classes. The three dimensional subspace of these dimensions thus corresponds to a cylinder with the stimuli arranged around the two rims. The next two dimensions seemed to indicate some sort of secondary rotation effect, as it pairs opposing stimuli, although the arrangement of these pairs is somewhat irregular. An undirected angular difference strategy (i.e. periodicity 180°) would require that pairs be arranged in a circle, whereas the g_3 strategy would not. The secondary rotation seemed identifiable by an examination of the plots of one of the primary rotation dimensions against one of the secondary rotation dimensions. The figure of eight patterns shown in Figures 7 & 10 are characteristic of the plots of $\sin \theta$ v $\sin 2\theta$, and $\cos \theta$ v $\sin 2\theta$. Similarily the other patterns corresponded to $\sin \theta$ v $\cos 2\theta$ and $\cos \theta$ v $\cos 2\theta$, thus identifying dimensions 3 and 5 as $\sin 2\theta$, $\cos 2\theta$ and the plane as a circle with parameter 2θ.

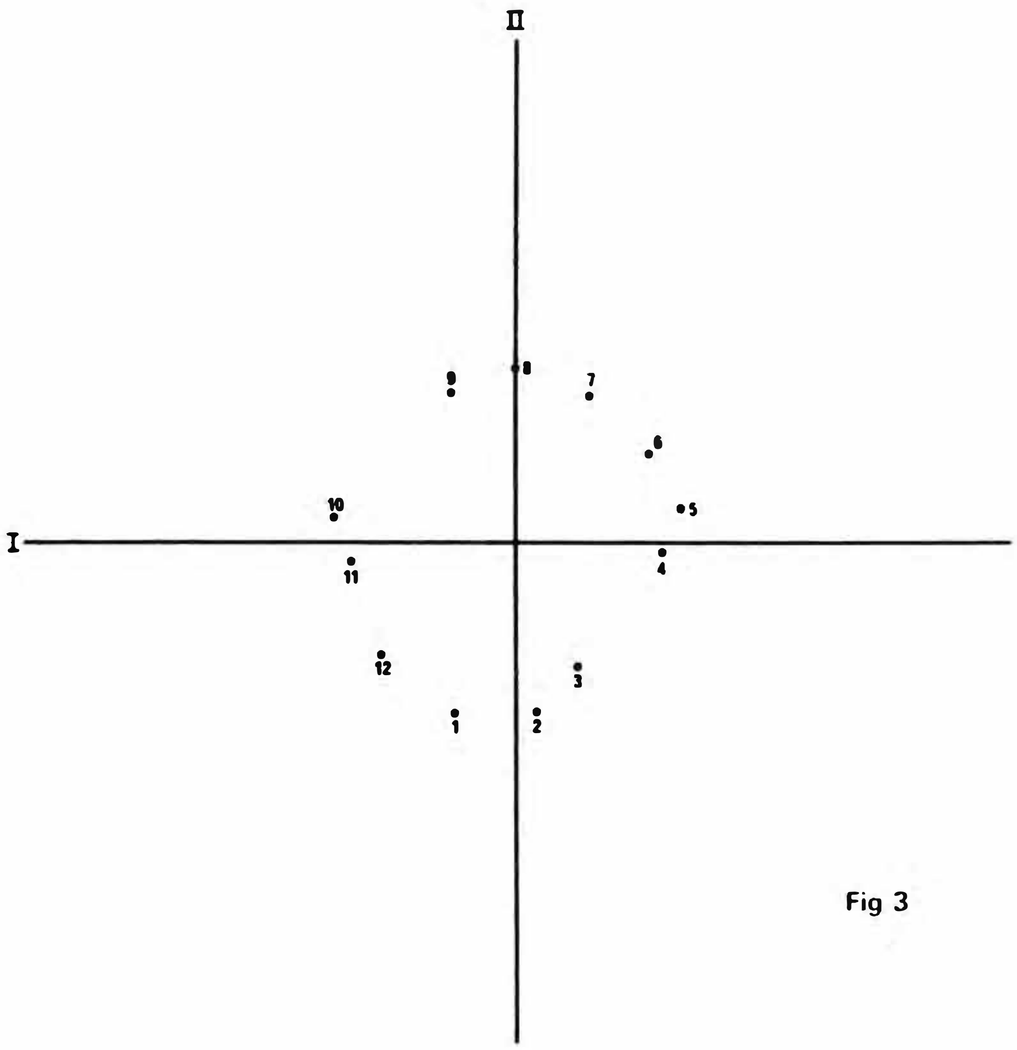


Fig 3

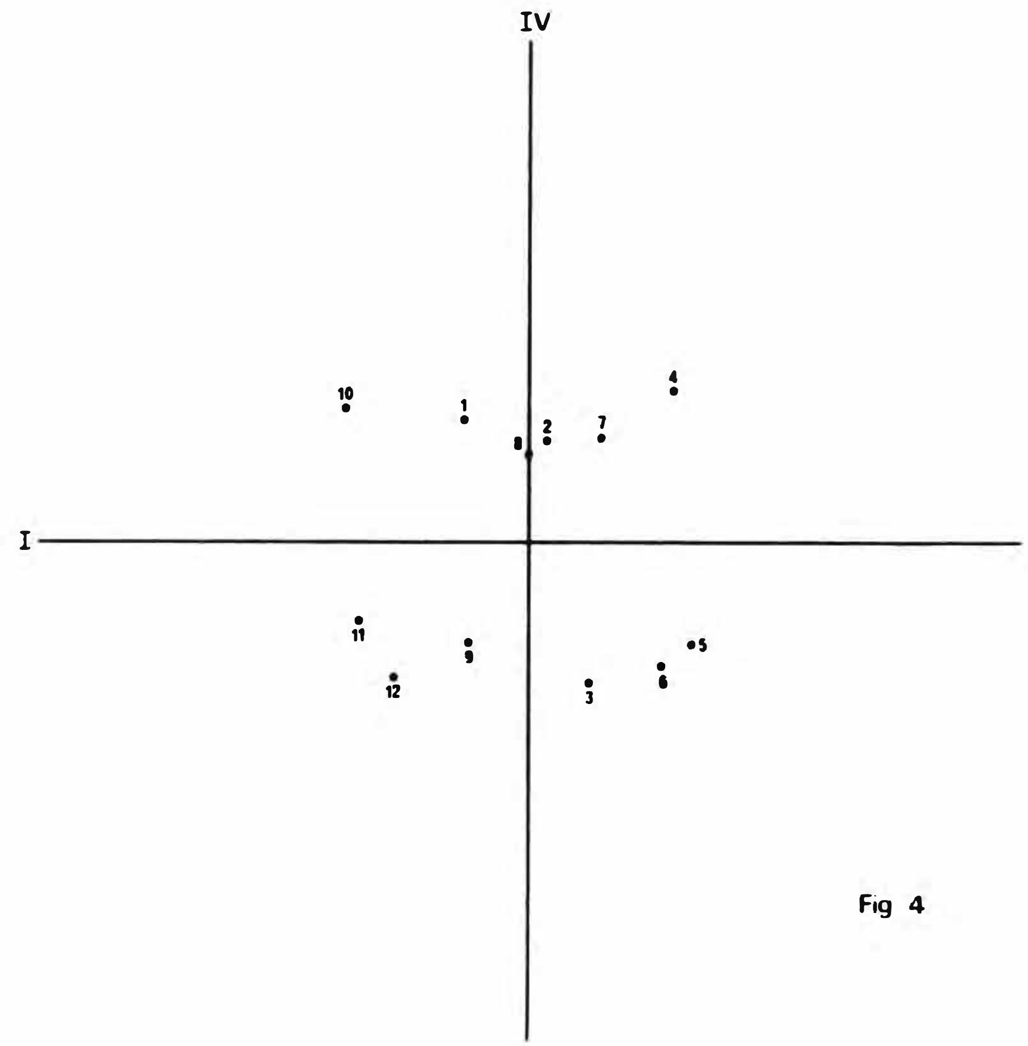
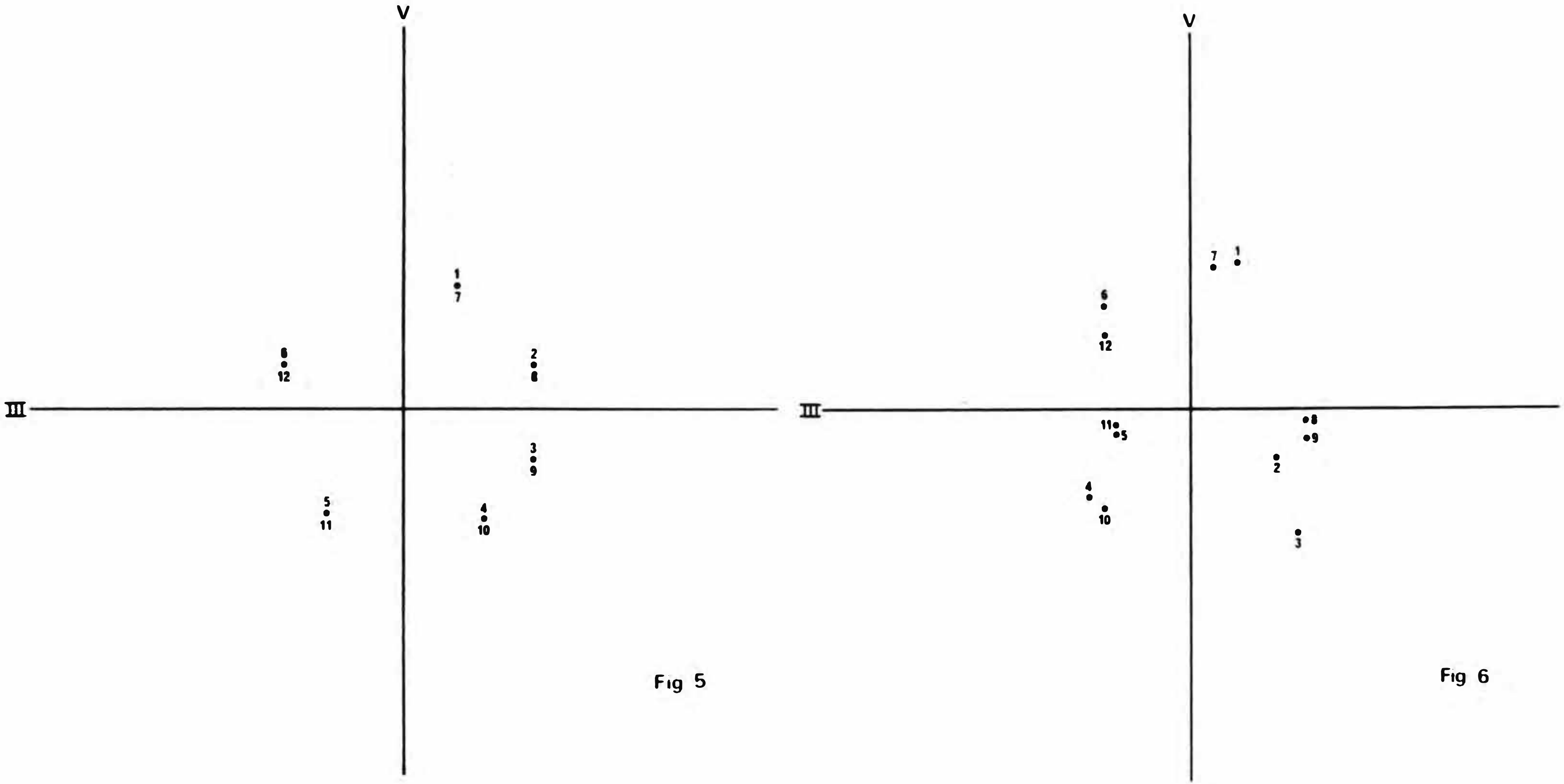


Fig 4

FIGURES 3-10: INDSCAL Scaling of Rotated Triangle Data: 5-D Solutions. The stimuli are numbered in order of increasing angle of rotation.



FIGURES 3-10: INDSCAL Scaling of Rotated Triangle Data: 5-D Solutions. The stimuli are numbered in order of increasing angle of rotation.

FIGURE 5 is an expected, not an obtained solution.

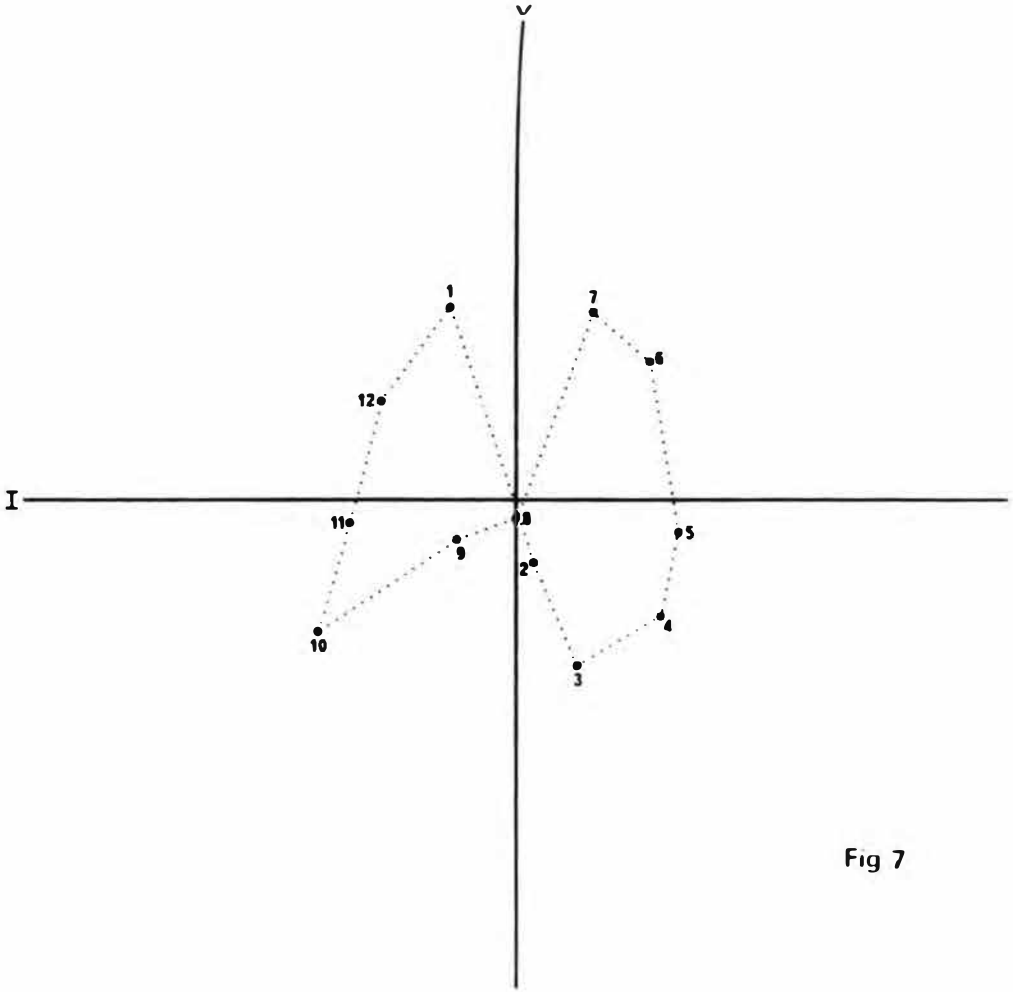


Fig 7

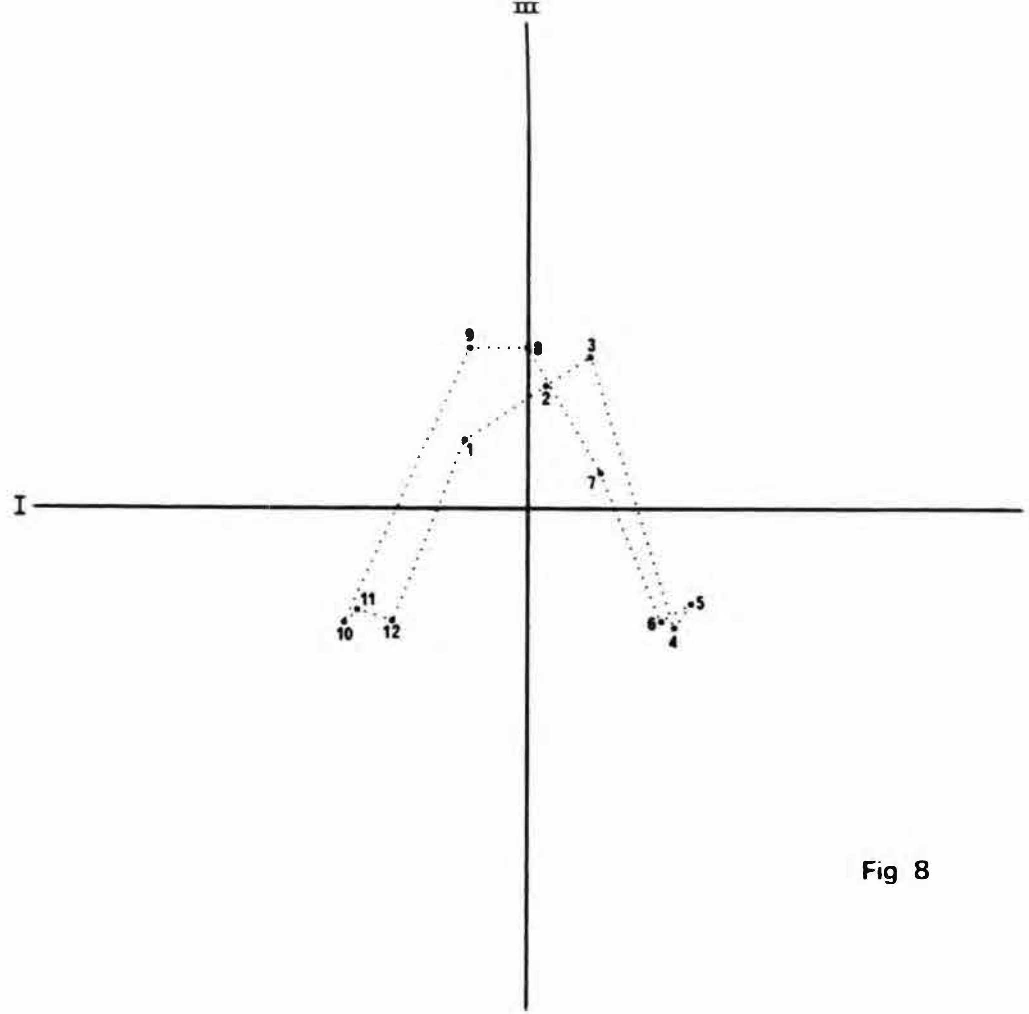


Fig 8

FIGURES 3-10: INDSCAL Scaling of Rotated Triangle Data: 5-D Solutions. The stimuli are numbered in order of increasing angle of rotation.

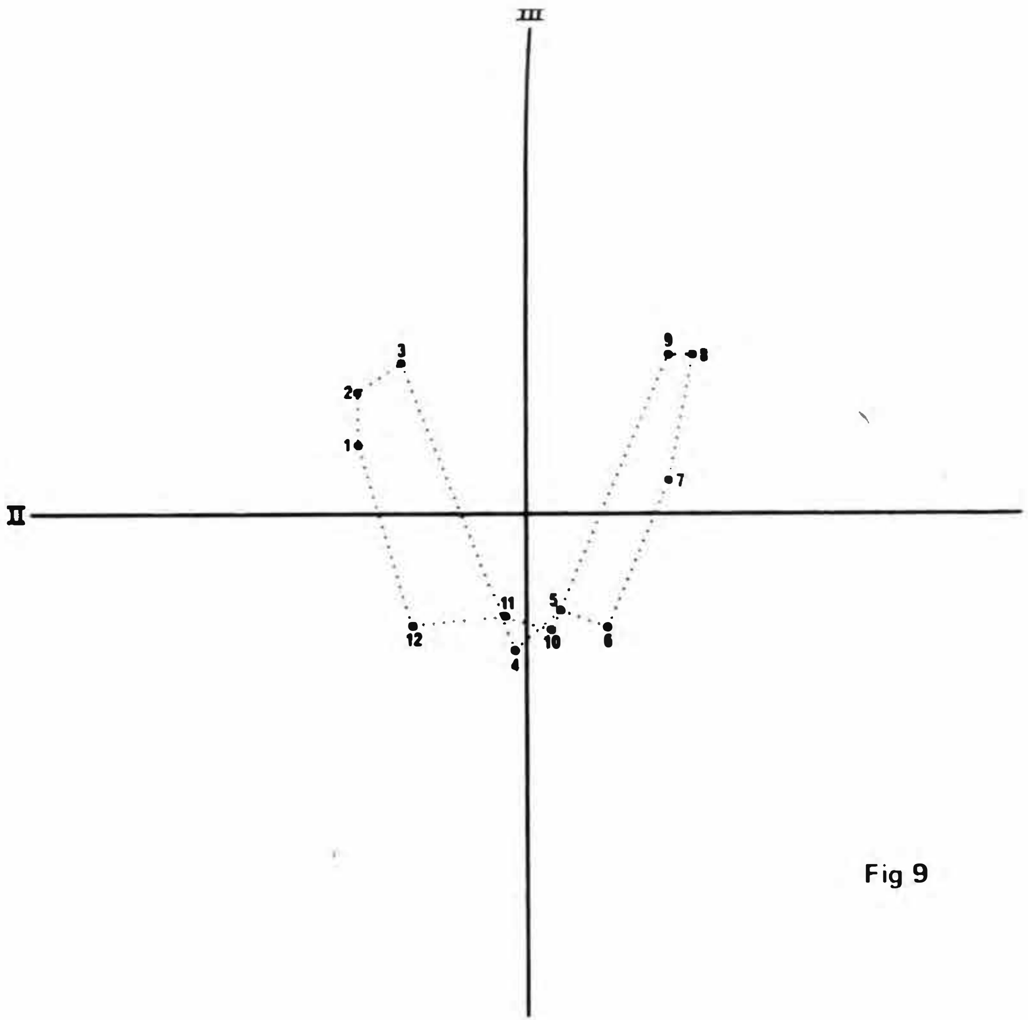


Fig 9

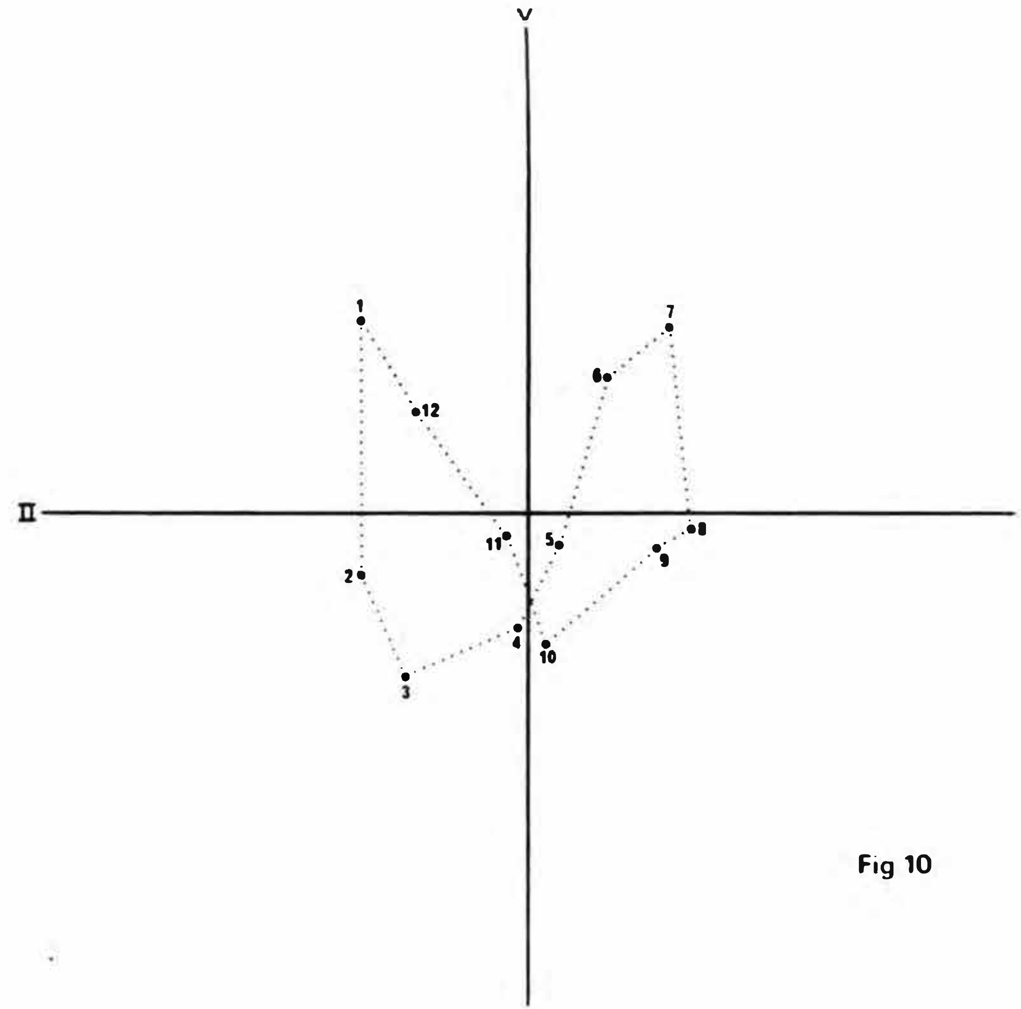


Fig 10

FIGURES 3-10: INDSCAL Scaling of Rotated Triangle Data: 5-D Solutions. The stimuli are numbered in order of increasing angle of rotation.

The group scaling configuration thus appears to provide an intuitively plausible representationof the structure of the similarities, in that it represents the transformations used togenerate the series and one of the two possible higher order strategies considered. However one would still need to show that the set of weightings on these dimensions does in fact give an accurate indication of individual subjects' strategies. Since we have only two physical dimensions, one of them two valued, upon which the subjective strategy must be predicated, we should be able to identify these by directly examining the similarity function i.e. by plotting the mean smilarity value for each angular difference, separately for g_2^0 and g_2^1. This requires assuming only that the measurement model is some function based on dimensional differences. As the main variations in the form of this function were expected to be in the periodicity of the group this was tested more quantitatively with a regression technique.

For each subject a series of six simple linear regressions of angular difference against similarity were performed, using three different periodicities (360°, 180°, 90°) to define angular difference, with separate analyses for each black/white equivalence class. Table 3 shows the significance levels attained in each regression, and also lists the dimension weights (averaged across the two dimensions for the circular dimensions) for each individual on the INDSCAL group configuration. In theory these allow for the salience of that dimension for that subject by altering the relative weightings assigned to dimensions in fitting distance to dissimilarity.

TABLE 3

INDSCAL Weights v Regression Significance Levels

SUBJECT	INDSCAL			REGRESSION ANALYSIS					
				g_2^0			g_2^1		
	g_1 (360°)	g_1 (180°)	g_2	g_1 (360°)	g_1 (180°)	g_1 (90°)	g_1 (360°)	g_1 (180°)	g_1 (90°)
1	.24	.34	.33		*				
2	.12	.00	.89						
3	.46	.31	.03	*		*	* *	*	
4	.62	.00	.10	* * *			* * *		
5	.62	.00	.10	* * *			* * *		
6	.63	.02	.03	* * *			* * *		
7	.37	.47	.15		* *		* *	*	
8	.47	.20	.06	*			* *		
9	.28	.41	.08		* *				
10	.26	.17	.70	* *					* *
11	.51	.00	.16	* *	*	* *	* *		
12	.48	.29	.06		* *		* * *		
13	.15	.54	.07		* *				* *
14	.60	-.07	.19	* * *			* * *		
15	.37	.27	.12	* *					
16	.34	.42	.16		* *		* *		
17	.51	.07	.22	* * *	*		* *		*

* $p < .05$

* * $p < .01$

* * * $p \ll .01$

The inspection of the similarity functions clearly indicated quite major diversions from those implied by the weighted Euclidean model. A typical result was quite a marked difference in the shape of the graph for g_2^0 and g_1^1, stimulus pairs. A few representative cases are described below.

Subject 1 The g_2^0 graph showed a U shaped function with a minimum at about 144°, indicating a primary rotation effect (periodicity 360°) with a fairly strong secondary rotation effect (180°). However the g_2^1 graph has a sawtooth pattern indicating either a periodicity 90° or some sort of 90° effect analag ous to g_3.

Subject 2 This graph was U shaped for g_2^1 and monotonically decreasing for g_2^0, indicating periodicities of 180° and 360° respectively.

Subject 3 Both graphs showed perfect monotone decreasing relationships indicating a primary rotation effect and little else.

So, in fact, apart from 5 subjects with pure strategies (primary rotation only) all subjects indicate strategies which clearly violate the assumptions of the Euclidean model, and in fact any model for dimensional representation (Beals, Krantz, Tversky, 1968), as the effect of differences on one dimension depend on the differences on another. A comparison of the two sides of Table 3 shows that apart from the 5 subjects using θ only (S4,S5,S6,S11,S14), there is very little correspondence between the dimension weightings assigned by INDSCAL and the strategies indicated by the regression significance levels. While of course they cannot be exactly equivalent the weightings do not even seem good approximations within the constraints of the independent dimensions model.

Discussion

The main point of interest to emerge from this analysis was that in spite of identifying that the individual subjective strategies utilized combination rules which in most cases repesented quite major violations of the conditions assumed by the Euclidean model, the INDSCAL procedure still successfully retrieved the major transformations which seemed to underly these strategies. The poor approximation provided by the weights is of reduced significance given the evidence which has subsequently been obtained that the weighted Euclidean model in general contributes little more than the unweighted model, and thus little significance can be attached to the weights. (See chapter III).

It will also be noted that the results of this analysis are somewhat inconsistent with those reported by Gregson, especially regarding the independence of g_1 and g_2. Gregson found that these were independent, a conclusion which would be more compatible with the INDSCAL configuration obtained here, whereas this study has found a substantial interaction. There are several possible reasons for this difference. It may be a genuine reflection of differences in the two sets of subjects involved. Gregson's data showed a much higher proportion of S's with maximum periodicity of 360^o for both g_2^o and g_2^1 (9 out of 16) and this becomes even higher after discarding subjects with insignificant regressions.

These subject differences may in turn have been a function of the differences in the experimental design regarding presentation of stimulus pairs. Gregson used a larger stimulus set (24) with an incomplete blocks design to select a subset of all the pairs for each subject. Each block consisted of all pairwise comparisons of a member from each of two subsets, each subset consisting of three stimuli of the same b/w class, in sequential orientation. Each subject was presented with four such blocks of 9 trials. It thus seems possible that this incomplete presentation and the narrow range of variation within each portion ofthe stimulus presentation series, could have a simplifying effect on the strategies employed by subjects.

Another possibility is that the use of pooled data on the periodicity of best fit, which in several cases differed for g_2^0 and g_2^1, may have obscured the effect of g_2^1 with single subjects. This pooling over periodicities thus requires the assumption of equivalent slopes across periodicities for each b/w class in order to test for a difference between classes.

We have thus in this case derived a configuration which represents plausible combinations of the objective transformations used to generate the series. However this does not necessarily indicate that these are the transformations which are combined in an idiosyncratic manner in individual subjects' cognitive strategies for generating the similarities. Neither analysis, INDSCAL or regression, would be sensitive enough for example to discriminate between the secondary rotation effect and g_3. INDSCAL must fit a dimensional representation, and a secondary rotation effect would be one best fitting solution even if g_3 was being used. With pooled data the secondary rotation solution might thus result even if most subjects with a similarity function peaking at 180° were using g_3 and only a few using 2θ. It could also result from exclusively g_3 strategies with some generalisation to stimuli near 180°. Similarly the coarse gradations between the stimuli make it difficult, by either direct inspection or regression, to distinguish between similarity functions continuously increasing towards 180° from those with a discontinuous jump at exactly 180°. The effect of the fairly small stimulus series involved has been exacerbated by the fact that each of the 10 possible angular differences is represented in only one of the g_2^1 or g_2^0 conditions. This produces only five points in the similarity/angular difference graphs for each subject (and b/w class) with which to identify the shape of the function.

This example thus reinforces the views discussed in the earlier chapters that the role of MDS procedures in identifying subjective perceptual strategies is clearly limited in either of the following conditions:-

(1) using data pooled over individuals without evidence that such a procedure is appropriate;

(2) relying on a post hoc dimensional interpretation to identify these strategies.

The blind application of MDS procedures in an exploratory mode does not provide a sound basis for reaching such strong conclusions. One needs to consider the extent to which alternative models can be accommodated by a given configuration (e.g. in terms of alternative dimensions or non-dimensional interpretations). One also needs to consider whether different configurations with quite different interpretations might also be found giving substantial equivalent degrees of approximation to the data. And finally, one needs to consider the possibility that a nonspatial model provides a better interpretation i.e. that one should not in fact have used MDS procedures.

The simple search for the solution of lowest residual error will do little to eliminate these other possibilities. As with all such statistical data reduction techniques the best solution must be found to the implicit model underlying a MDS procedure regardless of whether it is appropriate or not. The goodness of fit of this is not necessarily a good indication of the adequacy of the model as not all departures are equally significant in discriminating between it and other possible models. As previously commented MDS is particularly robust to departures from the assumed Euclidean model, indicating that quite significant departures may contribute relatively small amounts to the error function used by the model. In a similar manner slight alterations to the parameters of a given model might have quite significant theoretical importance but make a relatively small contribution to the error.

Thus one clearly needs to approach MDS interpretation from the point of view of comparing the a priori most plausible hypotheses, and focus ones attention on the aspects of the data which discriminate between these hypotheses. In the present situation this implies that one must attempt a more comprehensive and rigourous specification of the possible subjective strategies that might be involved with closed group transformations in general and spatial orientation in particular.

Spatial Orientation and Similarity

The orientation of objects in space plays, phenomenally, a considerable role in perception and it is therefore not surprising that attempts to characterise this aspect have been included in most theories of perceptual processes. Some of the earliest comprehensive treatments of the topic were provided in two quite similar theories by Sachs (1897) and Mach (1902). Both theories related perceptual similarity to the processes transforming one stimulus into another, Sachs attempting to account for this on a physiological level in terms of eye motions. The Sachs-Mach theory thus implies decreasing perceptual similarity against rotation up to a maximum dissimilarity at 180° rotation. They also considered the possibility of alternative more simple transformations at certain points. One specific instance for which they predicted increased similarity was when the two stimuli were related by a mirror reflection about the median plane. Sachs expected only the median plan to operate as an axis of reflection because of the symmetry of the oculomotor system at the physiological level with respect to the median plane. Mach considered the additional possibility of increased similarity at rotation differences of exactly 180° produced by attending to the coincidence of directions, disregarding the sense of direction. He did not however elaborate on the generality of this effect, for example whether it implied minimum similarity at 90°.

The Sachs-Mach theories thus consider some of the effects which correspond to the closed group transformations we have considered, including the possibility of different periodicities. It also however raised the possibility of some transformations corresponding to specific phenomenal effects particular to spatial orientation, such as mirror reflections.

A more comprehensive analysis of the phenomenal qualities of spatial orientation developed within the perspective of Gestalt psychology was provided by Goldmeier (1936/1972). Goldmeier's position was that similarity is a function of mental transformations and organizations, not of the physical properties of the stimulus pattern per se. Thus effects such as grouping, symmetry etc., contribute more than simply the number of parts in common. For Goldmeier these effects could be subsumed under the Gestalt principle of 'pragnanz'.

This he defined more precisely than in normal Gestalt theory usage to refer to a 'singular' or special value of a parameter. These are, in Goldmeier's terms, the exact values of a parameter that produce phenomenal qualities which are preferentially realised, for example equality of lengths, ratios, angles, etc. These singular values affect the similarity of pairs in a number of ways. First they are by definition particularly sensitive to distortion, and thus similarity decreases rapidly with only slight departures from singularities. Second the degree of singularity can function as a determinant of similarity, with singular figures resembling each other more than nonsingular figures. Third, the relation of two patterns to unpresented singular figures may affect their relationship to each other.

For spatial orientation the vertical and horizontal axes represent obvious singular values. Goldmeier describes a series of studies involving triadic similarities designed to show how the similarity of spatial patterns is affected by their relationship to the two singular axes. For example he presents several cases demonstrating that alignment to these axes can have a greater effect on similarity than rotation difference. Similarly mirror images are preferentially realised with respect to these axes. Goldmeier showed that, as predicted by Mach and Sachs, images reflected on the vertical are more similar than those reflected slightly off vertical. However he also found that images reflected on the horizontal axes, the worst case according to Mach-Sachs, were also more similar than axes between the vertical and the horizontal, although the effect was less than at the vertical.

These effects also occur between different parts of the same pattern. Thus two stimuli both symmetrical about either the vertical or horizontal axis are seen to be similar, much more than if the axis of symmetry is oblique. Again the effect of symmetry about the vertical axis dominated symmetry about the horizontal axis. Goldmeier offers a somewhat paradoxical explanation for this in that the up/down distinction is less confusable than left/right. More plausibly however this appears simply due to the experience with mirror reflections, because it is more natural to match image with reality by a turn in the horizontal than the vertical plane i.e. we maintain vertical constancy in preference to horizontal.

These singularities and the similarity determining strategies dependent upon them clearly pose a major problem for a dimensional representation. They imply that specific clusters of dimensional values can have an affect not determined by the orderings on specific dimensions, a clear violation of the independent dimensions assumption. It is thus clear that both the MDS and regression analysis applied in the present case would havelittle value as methods of identifying the cognitive strategies used by subjects to generate their similarities, as they would incorporate a clear bias towards those strategies for which an admissible representation exists in their underlying models.

The previous discussion has made it clear that there exists a wider range of possible strategies that could be involved in the present data than the ones previously considered. Thus what is required is a more comprehensive formulation of the plausible alternatives and a method of analysis which enables effective comparisons between them.

The main plausible alternatives for the given stimulus range are described below, together with a consideration of the implications of each strategy for a spatial representation. The possible strategies for a given individual should be some combination of the component strategies listed here.

<u>Strategy</u> 1 (Primary Ro tation)

Similarity decreases over increasing (minimum) angular difference. This would thus be represented as a circle in 2-D. The Euclidean distance function

$$d(\theta_1, \theta_2) = 2 \sin \left(\frac{\Delta\theta}{2} \right)$$

where $\Delta\theta = \theta_1 - \theta_2$

would monotonically decrease from a maximum at 0° to a minimum at 180° as required.

Strategy 2 (Secondary Rotation)

Similarity decreases over the minimum angle between undirected centre lines. This would be represented by a circle in 2-D, with points arranged according to 2θ. The Euclidean distance function

$d(\theta_1, \theta_2) = 2 \sin (\Delta \theta)$ would have maxima at 0° and 180°, and a minimum at 90° as required.

Strategy 3 (Black-White Reversal)

A constant reduction in similarity, independent of angular difference, is produced if the black side is reversed. This is represented by a two-valued dimension separating the equivalence classes of g_2^0.

Strategy 4 (Flip-over)

An increased similarity is produced at exactly 180° angular difference due to g_3 flipover transformations. This, like strategy 2, would cluster pairs of stimuli differing by exactly 180°, but would not require the pairs to be arranged in a circle. However Strategy 2 would also provide a solution for Strategy 4. An alternative more restrictive form of Strategy 4, which could operate only if the b/w orientation is preserved, has no cases existing in the present stimulus set.

Strategy 5 (Singular Directions)

An increased similarity is produced if both stimuli lie in singular directions, i.e. vertical or horizontal. We can subdivide this into the two cases of (i) one vertical with one horizontal, or (ii) both vertical or horizontal. 5(ii) will thus overlap with strategy 4.

One could also distinguish a third variation, an effect dependent on the singular relationship of the right angle. This would normally include 5 (i) as a subset, but in the present case would be indistinguishable from 5(i) as no stimulus pairs apart from those in singular directions differing by 90^{o}.

This strategy can be represented by a clustering of the four singular directions on one dimension. Alternatively it could be more economically approximated if Strategy 2 is also present, by utilising the theoretically unfillable space within the secondary rotation circle, i.e. by moving pairs (2,8) and (5,11) towards each other along the diagonal connecting them.

<u>Strategy 6</u> (Mirror Images)

Mirror images over the vertical or horizontal axis produce an increase in similarity. There are no exact cases of these in the present stimulus set, however, some approximate cases exist (i.e. of opposite b/w class and within 18^{o} of equal distance to vertical or horizontal axis). This strategy would have no low-dimensional representation as it is clearly not possible to construct any simple configuration in which distance is related to the presence or absence of mirror images.

<u>Strategy 7</u> (Black/white Rotation Differences)

A different rate of decrease of similarity with rotation exists depending on whether stimuli are in the same or different b/w equivalence class. Differences in both directions appear plausible. First, an increasing rate over increasing angles would imply an analogy to the well known results on the speed of mental rotation for same or different judgements (Shepard & Metzler (1971). A decreasing rate (of reduction) would imply that the b/w reversal effect has maximum effect at low angles, and diminishes as θ increases. If we express these strategies in terms of a penalty function governing the reduction in similarity when the pairs differ on g_2, then the two options would imply

(i) $P_1\ (g_2, \theta) = -\ (a\ b\ \theta + c)$

(ii) $P_2\ (g_2, \theta) = -\ (a\ d\ \theta + e)$

where $a = \begin{cases} 0 \mid g_2^1 \\ 1 \mid g_2^0 \end{cases}$, $b \geq 0$, $E(c) = 0$,

$d \leq 0$, $E(e) > 0$

Clearly there is no obvious dimensional solution for this class of function. This strategy is a refinement of strategy 3, and would become equal to that strategy only if in equation (ii), $d = 0$.

RESULTS - ANALYSIS TWO

(i) Regression Analysis

The first method of analysis used to attempt to identify this wider range of strategies was a multiple regression analysis. Seven variables were constructed corresponding to the predicted dimensional difference on each strategy for a given stimulus pair, as described in the descriptions of each strategy. Angular differences were expressed simply as multiples of an 18° rotation. The seven variables were thus defined as given below:

$X_1 = m$; $m = \min (K, 20-K)$

$X_2 = m'$; $m' = \min (2m, 20-2m)$

$X_7 = am$; $a=0 \mid g_2^0$, $1 \mid g_2^1$

$X_3 - X_6$ all defined as 0 or 1 depending on presence or absence of transformation defined.

The 7 x 66 matrix of scores on these variables for each stimulus pair combination constituted the set of predictor variables for a stepwise multiple linear regression against similarity, with separate analyses to be performed for each of the 17 subjects. Since the prediction equation is being used as a psychological model rather than simply for statistical prediction one needs to be mindful of possible effects caused by the order of inclusion of variables. If the variables are intercorrelated no simple strategy will guarantee the best combination of a given number of variables. In the present case however, the method of construction of the predictor variables ensures minimal intercorrelations, and thus a simple stepwise inclusion strategy was regarded as sufficient to select the set of significant variables.

Table 4 gives for each subject the list of significant variables, the corresponding partial correlation coefficients, and the multiple correlation for both the selected model and the full set of seven variables.

One clearly needs to exercise some caution in regarding these regression equations as representing accurate identifications of subjective strategies. With noisy data one could easily get capitilization on chance effects particularly with strategies involving a combination of effects. Several cases of this could be identified by the existence of implausible strategies such as negative mirror images, and these variables were deleted. However while Table 4 may not give exact indications for each individual, it should provide a reasonable guide to the range of strategy combinations used by this group of subjects, and the degree of consistency between them.

The regression equations appeared to identify six main clusters of subjects in terms of the types of strategies they employed. The first three groups relate to subjects whose strategies are restricted to those with a spatial representation (i.e. the INDSCAL dimensions). The first group contains five subjects identified as having pure primary rotation strategies, the second one subject using b/w reversal only, and the third two subjects using a primary and secondary rotation combination.

This however accounts for less than half of the total number of subjects. The remaining 9 subjects consist of those who violate the spatial assumptions by either (i) utilising strategies containing singularity effects or (ii) utilising non-independent combinations of the primary dimensions. Groups 4 and 5 contain 7 subjects with strategies containing singularities, and Groups 5 & 6 four subjects using BW effects, i.e. an effect dependent on the product of the Black/White reversal and angular differences. (Group 5 contains the two subjects with both types of inadmissible strategies).

TABLE 4

Multiple Regression Analysis Summary Table

Group No.	Subject No.	Full R(7)	Predictors	R	Partial Correlations
1	4	.877	θ	.872	-.872
1	5	.867	θ	.859	-.859
1	6	.888	θ	.882	-.882
1	11	.672	θ	.630	-.360
1	14	.865	θ	.840	-.840
2	2	.904	BW, (2θ)	.898	-.892, .384
			BW	.879	-.879
3	3	.687	θ, 2θ	.669	-.653, -.476
3	8	.749	θ, 2θ	.693	-.684, -.462
4	13	.548	2θ, SDR	.497	-.464, .302
4	7	.868	θ, 2θ, SDR	.844	-.751, -.814, .421
4	16	.717	θ, 2θ, SDO	.694	-.569, -.460, .328
4	9	.611	θ, 2θ, SDR, M	.588	-.335, -.539, .347, .300
			(-)		
4	12	.878	θ, SDR, SDO, (BW)	.876	-.838, -.613, .815, .489
			(-)		
			θ, SDR, SDO	.834	
5	1	.561	θ, BW, BWθ, SDO	.484	-.409, -.358, .357, .459
			(-)		
5	17	.755	θ, BW, BWθ, (Mirr)	.749	-.705, -.419, .410, -.272
			θ, BW, BWθ	.725	
6	10	.759	θ, BW, BWθ	.709	-.416, -.554, .269
6	15	.577	θ, 2θ, BW, BWθ	.527	-.502, -.355, -.319, .336

NOTES:
1. Full R(7) denotes multiple correlation between similarity and all seven predictors.
2. R denotes multiple correlation including only significant predictors as listed.
3. Partial correlations are listed in corresponding order by predictor list.
4. Minus sign in parentheses indicates predictor is significant in opposite direction to that hypothesised.
5. Predictors in paretheses indicate probable spurious correlations. Revised predictor list with this term dropped is shown on following line.

The most common singularity effects identified were singular directions, used by three subjects, and Flipover, used by another three subjects. One of this latter group also indicated a negative version of the singular direction effect, that is, a significantly reduced similarity at 90°. This seems to be most plausibly considered as a counter-effect to the Flipover effect, i.e. the exact opposite. That is this subject appears to be using a simplified discrete 2θ type deviation, that operates only at the 180° and 90° positions. One example of the Mirror-Image strategy was identified by the regression analysis.

The BW effects consistently supported the second of the two possible strategies considered here, as they all include a negative constant and positively increasing effect over θ. Thus the greatest effect of g_2 is at low angular differences, with the effect reducing as θ increases.

(ii) Rank Order Tests

The regression analyses represent far from ideal tests for the existence of singular strategies, since they depend heavily on linearity assumptions in the residuals after the more dominant continuous strategies have been partialled out. This procedure could thus be expected to be rather prone to Type I errors. An alternative procedure which would be less sensitive to errors of this type is to construct tests based on ordinal comparisons between specific stimulus pair categories.

The Singular Directions strategy for example predicts a peak in the similarity function at 90°, up to which both θ and 2θ strategies expect similarity to decrease. Thus the existence of Singular Dimensions, in conjunction with either or both angular difference strategies, should produce the following orderings on the similarity of pairs with these angular differences

(i) $S(90^{\circ}) > S(72^{\circ})$ and

(ii) $S(54^{\circ}) > S(72^{\circ})$

The existence of b/w reversal effects is a complicating factor, in that in the present design each angular difference always occurs with the same value of g_2. However since the values are g_2^1, g_2^0, g_2^1 for 90°, 72°, 54° respectively this will increase the probability of Type II errors only. The orderings above were tested using a Mann-Whitney U test on the two groups of pairs with the corresponding angular differences. Table 5 lists the minimum U values, the two-tailed probability of this, the obtained and predicted directions of the difference (i.e. the group with the highest similarity). Obtained values are listed only for differences exceeding one standard deviation, predicted values are listed if the combination of strategies identified by the regression analysis produces an unambiguous prediction.

Table 5 shows only one subject (S13) with the required $S(90^{\circ}) > S(72^{\circ})$ relation. This result is significant ($p = .035$) for a one-tailed test, S13 being one of the 3 subjects identified as using Singular Directions by the regression analysis. Neither of the other two subjects (S7, S9) showed even a trend towards a 90° effect. S13 did not show even a trend for the second comparison, which corresponds with the regression analysis, where this was the only subject not to show a significant primary rotation effect (apart from the S using BW only).

TABLE 5

Rank Order Tests for Singular Strategies: Singular Directions

Subject	90^{o} v 72^{o}				54^{o} v 72^{o}			
	$U_{(4,8)}$	Signif.	Obtained Highest	Predicted Highest	$U_{(8,10)}$	Signif.	Obtained Highest	Predicted Highest
1	15	-	-	72^{o}	22.5	.13	72^{o}	-
2	4	.048	72^{o}	72^{o}	0	.000	72^{o}	72^{o}
3	12	-	-	72^{o}	16.5	.038	54^{o}	54^{o}
4	6.5	-	72^{o}	72^{o}	11	.008	54^{o}	54^{o}
5	6	.11	72^{o}	72^{o}	30	-	-	54^{o}
6	10	-	72^{o}	72^{o}	12	.012	54^{o}	54^{o}
7	13	-	-	90^{o}	21.5	.11	54^{o}	54^{o}
8	15.5	-	-	72^{o}	14	.021	54^{o}	54^{o}
9	11	-	-	90^{o}	26	-	54^{o}	54^{o}
10	4	.048	72^{o}	72^{o}	5.5	.001	72^{o}	-
11	13.5	-	-	72^{o}	30	-	-	54^{o}
12	16	-	-	72^{o}	2.5	.000	54^{o}	-
13	5	.073	90^{o}	90^{o}	29	-	-	54^{o}
14	16	-	-	72^{o}	34	-	-	54^{o}
15	7	-	72^{o}	72^{o}	36	-	-	-
16	14.5	-	-	72^{o}	36	-	-	54^{o}
17	8	-	72^{o}	72^{o}	19	.068	72^{o}	-

NOTES:
1. Predicted Highest based on strategies identified by multiple regression analysis.
2. Significance levels only reported for values of U at least one standard deviation from expected value.
3. Significance levels are two-tailed probabilities.

Another feature of note in Table 5 is the greater strength of the rotation effects between 54^o and 72^o than between 72^o and 90^o. It indicates 5 significant and 2 nonsignificant trends for the former comparison, compared with 2 significant and 5 nonsignificant trends for the latter. This appears to imply that the shape of the similarity function over angular difference is negatively accelerating.

A further point is the existence of two significant (and two nonsignificant) trends for $S(72^o) > S(54^o)$. Since this comparison pits the B/W effect against the rotation effects this clearly indicates a strong g_2 component in the strategies of these subjects. These subjects (S1, S2, S10, S17) are four of the five subjects identified by the regression analysis as using BW related strategies.

A similar procedure was adopted to test for the existence of Mirror Image strategies. The group of approximate mirror image pairs was tested against a group containing the same angular differences. This identified three subjects with trends bordering on significance. These were

S9 $(U_{8,12} = 27, p < .05)$

S17 $(U_{8,12} = 27.5, p < .1)$

S13 $(U_{8,12} = 28.5, p < .1)$

(all probabilities are one tailed).

The result for S17 however was in the opposite direction to that predicted, implying a reduced similarity for mirror images. This was also identified in the regression analysis but was deleted as being an implausible strategy. S9 was the only subject identified in the regression analysis as utilising a positive mirror image strategy.

TABLE 6

Rank Order Tests for Singular Strategies: Mirror Images

Subject No.	$U_{(33,33)}$	Significance Level
1	435	.17
2	69	.00
3	412	.09
4	427	.14
5	479	.43
6	431	.16
7	538	.98
8	464	.42
9	474	.38
10	161	.00
11	505	.65
12	483	.46
13	485	.48
14	531	.91
15	487	.49
16	539	.99
17	534	.94

(iii) Constrained Multidimensional Scaling

This analysis employed the technique of constrained multidimensional scaling to test the hypotheses that individual subject strategies were as identified in the previous analyses. This procedure is of course limited by the fact that only strategies corresponding to a spatial representation can be tested in this manner. That is the only hypotheses that can be directly tested are strategies corresponding to the three INDSCAL dimensions. The existence of inadmissible strategies can only be inferred in a negative fashion by the poor performance of the admissible strategies.

Six subjects were selected for this analysis to represent the subject clusters identified by the regression analysis. Subjects were chosen according to the highest multiple R from the regression analysis, hopefully an indication of less noisy data.

The scaling program used was CONSCAL, an unpublished but highly efficient algorithm developed by Noma and Johnson, (1979) and run on the University of Michigan Amdahl 470V/7 computer. This is a nonmetric program which alternates between minimisation phases based on restrictions implied by the data and by the constraints until convergence is obtained.

Separate analyses were carried out on each of the six selected similarity matrices. A series of unconstrained scalings were used to test for the appropriate dimensionality, with both random and hypothetical starting configurations being used to minimise distortions caused by local minima. Then the most plausible constraints were tested based on the strategies predicted from the previous analysis and the indications of appropriate dimensionality. Additional scalings were performed if the expected options failed to produce satisfactory results.

The hypotheses that subjects were using either or both of the rotation strategies were tested by applying constraints on the polar co-ordinates of the stimuli in a two-dimensional subspace. The hypothetical configurations were thus specified as equal radial distances, and angles ordered as given by the primary and secondary angular differences from a fixed starting point. For primary rotation both the radial distances and angles were constrained under strong monotonicity, implying that the tied radii were to remain equal, and the angles were to remain strictly ordered. Thus the only degrees of freedom remaining in this plane were alterations in the size of the intervals separating stimuli on the circumference of a circle. This therefore corresponds to a strongly ordered scaling in an effective dimensionality of one.

For the secondary rotation dimensions, strong monotonicity was again applied to angular differences but only weak monotonicity was applied to distances. Since all distances were specified to be equal this effectively means no constraints were applied to distances at points from the origin, only to the orderings of the angles of the vectors from the origin. The increased degrees of freedom allowed in this plane were to accommodate possible significant distortions corresponding to singular strategies.

Table 7 shows the final stress levels after either convergence to a locally minimum solution, or a maximum of 25 iterations were exceeded for all conditions under which a solution was sought for each of the six subjects.

TABLE 7

CONSCAL Stress Levels: Summary Table

Scaling Conditions	Subject Number				
	6	7	8	10	17
<u>Unconstrained</u>					
2D - RS	.066	.215	.189	.199	.240
- BFS	.066				
3D - RS				.107	
- BFS				.102	
4D - RS		.088		.041	.069
- BFS	.026	.053	.078		
<u>Contrained</u>					
θ, 2D	.092; .089		.249		
θ, 2θ, 4D		.062; .057	.183		
θ, 4D			.107		.094
θ, 2θ, 5D			.102		
θ, BW, 3D				.145	
θ, BW, 4D				.089	.105
θ, BW, 5D					.052

<u>NOTES</u>: 1. Multiple entries indicate more than one scaling.
2. Abbreviations Used:

- 2D : 2 Dimensional Solution
- RS : Random Start
- BFS : Best Fit Start
- θ : Points Constrained to Lie on Circle in Order of θ
- 2θ : Points Constrained to Lie on Circle in Order of 2θ
- BW : Dimension Projections Constrained by Black/White Code Value

<u>Subject 6:</u> This subject was selected as a representative of the primary rotation only group. As shown in Table 7 the unconstrained solution shows a good fit in 2-D, and only a slight reduction in fit when this solution is constrained to lie on the primary rotation circle.

<u>Subject 7:</u> This subject was selected to represent the group including θ, 2θ and singularity effects. However an excellent fit was obtained using a solution in 4D constrained by θ and 2θ.

<u>Subject 8:</u> This subject was initially identified as using a θ and 2θ only strategy combination. However this combination proved to be a very poor estimation of what seems to be a 4-D solution. When constrained in θ only the remaining 2D showed an approximate 2θ solution with a consistent distortion in the relative orientation of the two circles (See Figure 11). This appears to line up the pairs (7,12), (2,8), (4,10), and the quadruple (1,5,6,11). This appears to suggest an approximate flip-over operation i.e. a similarity between pairs which can be related by the g_3 operation described by Gregson. While there are no exact cases the pairs (7, 12), (1, 6), and (3, 8) represent approximations with 18°. The alignments (2, 8), (5, 11) correspond to the 180° effects along the singular dimensions, and the pair (4, 10) are opposites in a non-singular direction. Thus it appears to be a 2θ strategy which includes g_3 which has produced this systematic distortion. The distortion could be a result of relative size of the effect at different points, which could be due to a chance relative lateral positioning of the stimuli

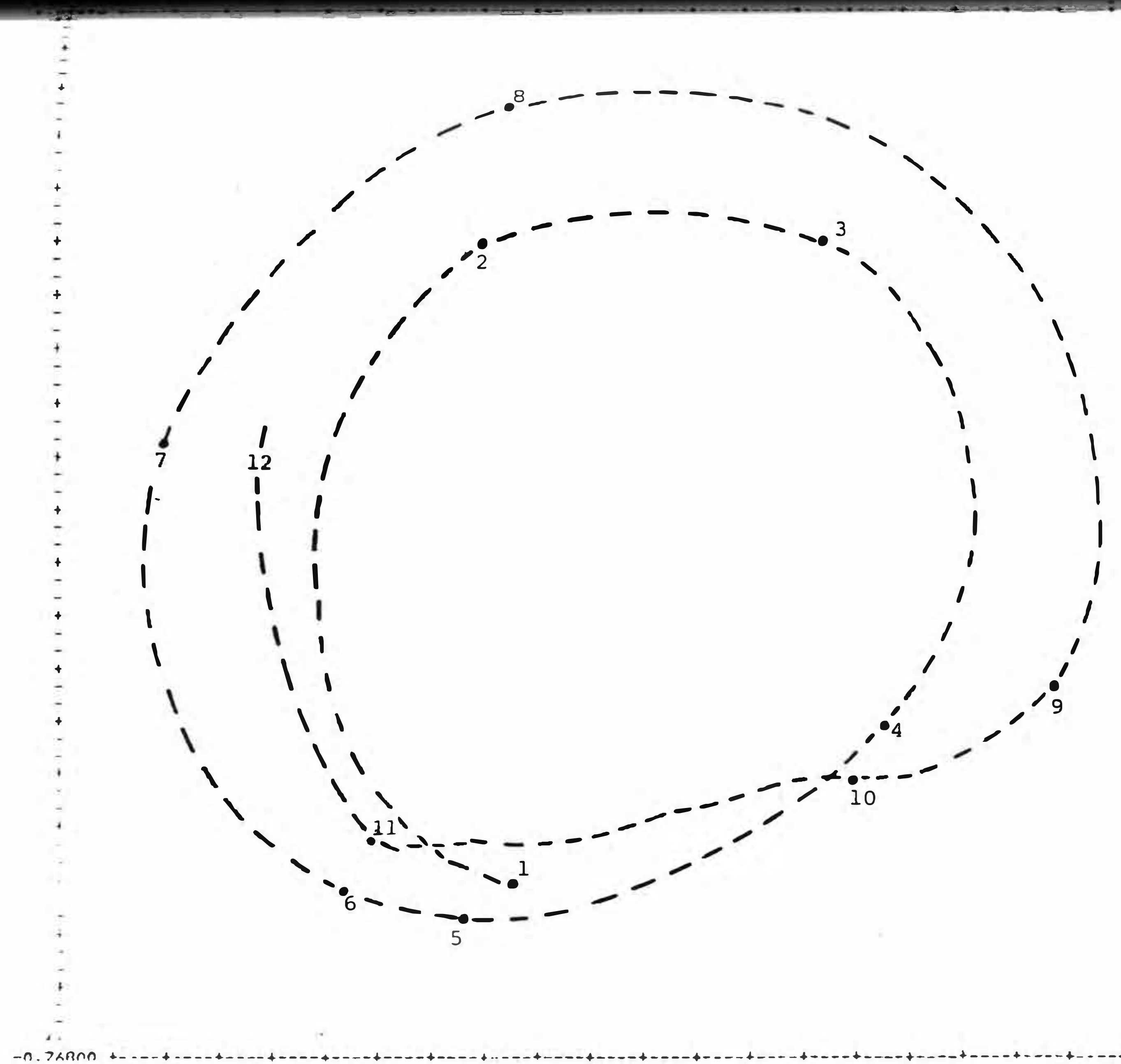


FIGURE 11: CONSCAL Solution, Subject 8, Dimension 3 v Dimension 4 of 4-D Solution, Dimensions 1 and 2 constrained by θ.

in these pairs (as the effect could be expected to be dependent on this).

Subjects 10 and 17: These subjects represented the group including b/w related strategies. In both cases the best fit appeared to be in 4-D. Including the constraints of θ and b/w appear to have resulted in fairly limited increases in stress, although these by themselves do not produce a satisfactory solution. S10 appears to require one additional dimension, S17 at least one, possible two. However no simple interpretation could be given to any of these additional dimensions. They most likely simply represent approximations to the most dominant components of the non-spatial b/w related angular difference strategies. The only method of pursuing this further within a spatial methodology would be to examine the structure of the residuals after the admissible dimensions have been extracted. This would however be virtually equivalent to the regression analysis previously considered, except with the angular differences monotonically transformed into direct Euclidean distances across the circle.

The previous analyses have suggested that there is insufficient reliability in the metric properties of the data for this option to be pursued.

Discussion

The series of analyses described above represent three different methods of testing specific representations for the matrix of similarity measures generated by an individual subject. The three procedures emphasise linear, ordinal, and spatially representable relations respectively to distinguish between the competing hypotheses.

The presence of some inconsistent results between the different methods indicates that the data are fairly noisy with respect to the finer grain distinctions required to distinguish between the various possible hypothetical strategies. The strong linear assumptions implied by the regression analysis in particular have caused some spurious identifications, although most results are in substantial agreement with the other procedures. The precise identification of an individual strategy with tests based on weaker ordinal constraints however would require a much more extensive stimulus design than was able to be utilised here.

Thus the evidence from the present study must be regarded as somewhat inconclusive regarding the exact identification of each individual strategy. It does however provide an indication to the extent that the total set of subjects can be accomodated by strategies of a particular type .

It is clear that the continuous dimensional strategies based on primary and secondary angular differences do provide a good broad level approximation of the similarities of most stimulus pairs for most subjects. There is some evidence that some subjects are using singular strategies as variations from these at a more fine grain level, although the extent of this is somewhat uncertain.

While each distinct singularity effect appeared reliably identifiable in only one or two subjects the overall indications were that between a quarter and a half of the subjects were using strategies that depended to some extent on the specific vertical and horizontal orientations.

The evidence is much clearer that departures from the spatially admissible strategies i.e. those corresponding to independent contributions of differences in the INDSCAL derived dimensions, did occur for subjects whose strategies involved a substantial g_2 component. While only about 30% of subjects were in this category, for all of them (except for the one subject using g_2 only) the size of the g_2 effect was clearly dependent on θ.

CHAPTER VI

SEMANTIC STRUCTURE OF EMOTION LABELLING

The present study represents another demonstration of the application of a theory-testing orientation to the interpretation of MDS configurations. This study will consider the fairly infrequently researched area of emotion labelling.

The theoretical developments which will be drawn upon to guide the reseach will include both that relating to the empirical structure of emotion itself, and also the more general linguistic considerations relating to the use of semantic labels rather than the stimulus objects themselves. As essentially private experiences emotions are clearly more susceptibre to linguistic ambiguities than physically realizable objects.

While most studies of emotion have used emotion words, some have attempted to study the structure of the emotional state more directly by the use of facial expressions as the stimulus objects (e.g. Osgood 1966; Schlosberg 1952, 1954). However as Fillenbaum and Rapo port (1971, p.102), point out an isomorphism between the structures of expressed and labelled emotion would be nice but by no means guaranteed or even conceptually necessary. The language of emotion clearly covers a much wider domain than simply effective feeling. This point is also made by Miller and Johnson-Laird (1976) in their classic review of the role of language in perception. They argue that this distinction has often been overlooked in many previous studies of this area. The language of emotion in fact provides for a distinction between an emotion, the cause of an emotion and the object of an emotion.

Miller and Johnson-Laird argue that it is the semantic structure rather than that of the pure experience that acts as the main determinant of behaviour. While they do not attempt any extensive analysis of the semantic structure of emotion labelling they endorse this as an important if imperfectly understood aspect of human behaviour. They criticise information processing systems which do not provide any adequate way of representing the emotional forces which are such an important part of human life.

Dimensions of Emotion

A wide variety of conceptual schemes have been provided as possible bases for the classification of emotional phenomena. The problem was first considered by Wundt (1897) with other important contributions including Harlow and Stagner (1933), Schlosberg (1954) Block (1957) Arnold (1960) Schachter et al (Schachter and Wheeler, 1962; Schachter and Singer, 1962) and Young (1961). While there are some differences in these approaches many of them can be resolved in terms of varying emphases on the different aspects of emotional meaning referred to above.

An extensive review of all these schemes by Davitz (1969) showed substantial agreement on three major dimensions of emotion, denoted activation, relatedness and hedonic tone. A dimension of activation or affective intensity is mentioned by almost every writer, although under a variety of labels. For example Wundt in his tri-dimensional theory of feeling specifies excitement-quiescence as one major dimension. However as Arnold (1960) comments a dimension of intensity is common to most psychological activity and therefore not an important distinguishing factor.

The term relatedness refers to the individuals tendency to approach or avoid the aspect in the environment which is the source of the emotional experience. A related concept considered by some theorists, notably Schachter, is that of competence. While Schachter did not specifically define this as a separate dimension his studies of emotion clearly demonstrated the importance of cognitive-social factors in the definition of emotional states. That is, the individual's perception of himself and the social situation is an important determinant of the way he interprets his emotional feelings. The term competence was suggested by Davitz (1969) as the most appropriate term for the dimension corresponding to variations in this regard.

Hedonic tone is the third primary dimension of emotion mentioned by almost every theorist, usually referred to as a bipolar pleasantness-unpleasantness scale. Thus there seems a reasonable degree of consensus on these three (or four) scales as primary dimensions of emotional meaning.

The attempts to match these theoretical dimensions with empirical structural representations have been only partially successful. Osgood (1966) used factor analyses of facial expressions to identify three factors which he termed Pleasantness, Activation and Control, although the factor structure seemed more to suggest a number of distinct clusters rather than three continuous dimensions. In another fairly comprehensive study by Davitz (1969) 50 emo tional states were described using a protocol of 556 possible statements. A cluster analysis performed on the basis of the overlap between the statements applied to different emotions produced 12 clusters. These were interpreted as three different levels on the four primary dimensions distinguished above. Rather than a simple ordering on these scales Davitz interprets them as a partial order of one positive and two types of negative states. These are denoted hyperactivation and hypoactivation corresponding to active and passive type negative reactions.

Fillenbaum and Rapoport used a variety of formats to generate similarity data for a set of 15 emo tion names using a combination of MDS and clustering methodology. They found no clear interpretation to the MDS configuration although the ubiquitous pleasantness, unpleasantness scale was in evidence. The clustering solution also did not offer any obvious interpretation that did not present several inconsistencies, and the clusters were by no means compact. They conclude (p.123) that a spatial model does not appear appropriate in this instance.

Thus while there is sufficient support for the conceptual schemes for the semantic structure of emotion discussed here for them to be considered as plausible approaches, no adequate representation has been found for them in empirical structure. While this may mean as Fillenbaum and Rapoport suggest that a spatial representation is inappropriate, these previous evaluations have shown insufficient conceptual clarity and rigour for other conclusions to be eliminated. Most previous studies have simply attempted to identify loosely defined dimensions or types in a fairly simple dimensional or clustering representation. One clearly needs to formulate the conceptual system more precisely in terms of a theory which leads to precise predictions of empirical structure. Davitz's typal scheme for example seems to involve some confusion between dimensions and values on dimensions and does not adequately represent the expected relationships among the types.

The purpose of the present study is thus to attempt to formulate a more adequate theoretical structure to organise this aspect of the semantic domain, and to develop an appropriate methodology by which one can evaluate it.

The combination of dimensional and typal information contained in the classification schemes suggests that Guttman's facet theory approach might represent an appropriate methodology to employ here. The theory to be evaluated was derived to be consistent both with the theoretical considerations discussed above, and the apparent structure obtained in the present study, and thus no claim is made for this to be an independent evaluation of a genuinely prior hypothesis. The intention is rather at this stage limited to suggesting a possible line for conceptual development in this area, and a methodology by which structural constraints implied by such schemes can be empirically evaluated.

Method

A set of 16 emotion labels was chosen following largely on those used by Fillenbaum and Rapoport (1971). While a more systematic method of ensuring a representative selection would have been preferred none of the classification schemes described above were considered sufficiently well defined for this. The standard complete presentation paired comparisons design with pairs in randomised order and balanced for order within pairs, was again selected for this study. While this offers the advantage of methodological simplicity at this early stage in this area of investigation some method enabling a substantially larger stimulus set size would clearly ultimately be required for any comprehensive evaluation of such a wide empirical domain.

Stimulus pairs were presented in the form of a written questionnaire, each page containing two columns of 15 pairs with a space beside each pair to record the chosen response from an 11 point magnitude estimation scale. This format enabled subjects to complete the 120 pairs in 20-30 minutes. The questionnaire was administered to a group of 22 first year psychology students at the Gippsland Institute of Advanced Education. The instructions printed on the front cover of the questionnaire requested the subjects to assess the similarity of the emotions referred to by each pair of words on a scale of 0 to 10. They were expressed in these general terms in order to encompass the whole semantic structure of emotional definition, without focussing attention on any specific aspect such as the nature of the feelings accompanying emotional states, the situation they occur in, objects or causes of emotions etc. Appendix A, which contains a copy of the questionnaire, lists the actual instructions used.

The sample size was reduced to 11 by discarding subjects with a substantial proportion of zeros in their similarity matrix. While not an inadmissible response a large proportion of zero similarity assessments indicates that the response scale has been truncated at one end, with no distinction made between distances over a certain size.

The 11 similarity matrices were then analysed by ALSCAL (Takane et al 1977), after first being converted to dissimilarities, the preferred format for this procedure, by reversing the scale. As previously discussed the ALSCAL program does not allow for any direct test of the appropriateness of its basic subjective metrics assumption. However Young has commented (personal communication) that some indication of this can be obtained by comparing the fit measures obtained under a number of options that place differing emphases on this assumption. For example by varying the conditionality option from unconditional (i.e. treating the subjects simply as replications) to matrix conditional additional degrees of freedom are provided for the subject weights to account for the variance in the data. Conversely a move from metric to ordinal scaling would place less emphasis on variations in subject weights, as an alternative source of freedom would be available to account for the data.

Results

Ordinal matrix conditional scaling were carried out first to determine the dimensionality required to represent the pooled data under the most unrestricted model. Table 8 shows both goodness of fit criteria generated by this procedure, for each dimensionality from 2 to 4. The Stress value shown is the average (RMS) individual stress using Kruskals Stress

TABLE 8

Average Goodness of Fit by Dimensionality
Weighted Ordinal Solution

Dimensionality	Stress	RSQ
2	.289	.516
3	.200	.599
4	.180	.444

Formula 1 (1964). The RSQ value is the average squared data-distance correlation.

The 4-dimensional solution appears somewhat sub-optimal, with quite large differences in individual stress values, producing the low mean RSQ value. However a more extensive search for a solution in a dimensionality as high as this would normally be justified only if three dimensions appeared clearly insufficient. The 3-dimensional solution in this case shows an adequate if not exceptional fit, but more importantly to offer a regular systematic interpretation in terms of the expected theoretical structure.

Table 9 shows the goodness of fit figures under four different scaling conditions; individual solutions, weighted ordinal, replicated ordinal and replicated metric. This shows that the main source in differences is in the allowances for different metrics. The ordinal solutions are substantially better than the metric solution. There is little difference between the weighted and replicated models, indicating that no significant contribution to fit for any individual is provided by variations in dimensions weights.

A comparison of individual and group scaling stress values shows that a few subjects (S3, S4, S5, S6) are not very well fitted by any of the group solutions. This suggests that there is not complete homogeneity of the semantic structure of this group of subjects.

The 3-D weighted ordinal solution however appears to represent an adequate if not exceptional representation of the majority of subjects and thus this is the scaling for which an interpretation will be attempted.

TABLE 9

Individual and Average Goodness of Fit - 3-Dimensional Solutions

Subject	Individual Solutions		Replicated Ordinal		Weighted Ordinal	
	Stress	RSQ	Stress	RSQ	Stress	RSQ
1	.117	.885	.173	.695	.166	.723
2	.125	.844	.192	.620	.185	.653
3	.134	.813	.229	.462	.234	.453
4	.110	.896	.228	.469	.232	.457
5	.124	.838	.234	.443	.233	.455
6	.122	.863	.221	.498	.231	.463
7			.177	.675	.168	.732
8	.134	.856	.191	.626	.200	.605
9			.195	.610	.187	.661
10			.189	.631	.187	.642
11			.169	.702	.159	.742
OVERALL			.201	.585	.200	.599

NOTE: Not all individual solutions produced.

The two 2-D subplots of this solution are shown in Figs 13 and 14.

The most notable thing about this structure is the fact that almost all the stimuli are located fairly equidistant from the origin. That is, the structural information appears to be represented as a spherical surface rather than a series of dimensions. This also suggests that an appropriate interpretation scheme could be provided by Guttman's facet theory scheme for relating ordered regions to qualitative theoretical facets. The present structure is a three dimensional extension of a radex with a series of concentric spheres, rather than circles. The dotted lines drawn on the graph subdivide the circular order into a series of wedge-shaped regions that appear both spatially and conceptually distinct. The order implied by the contiguity of these regions is in fact only slightly more complex than a 2-dimensional radex, as can be seen from the plot of the 2-dimensional solution.

While dimensionality and the number of facets are not related in any simple way this gives some indication of the complexity of the formal structure of the conceptual system. Although it depends on the number of values within facets the number of facets is usually much higher than the number of dimensions (see Borg 1977a for a discussion of this point). The small number of distinct regions indicates that even with the minimum two values per facet no more than three or four would be required to produce the circumplicial order, with another to produce the lineal orderings (along the radii).

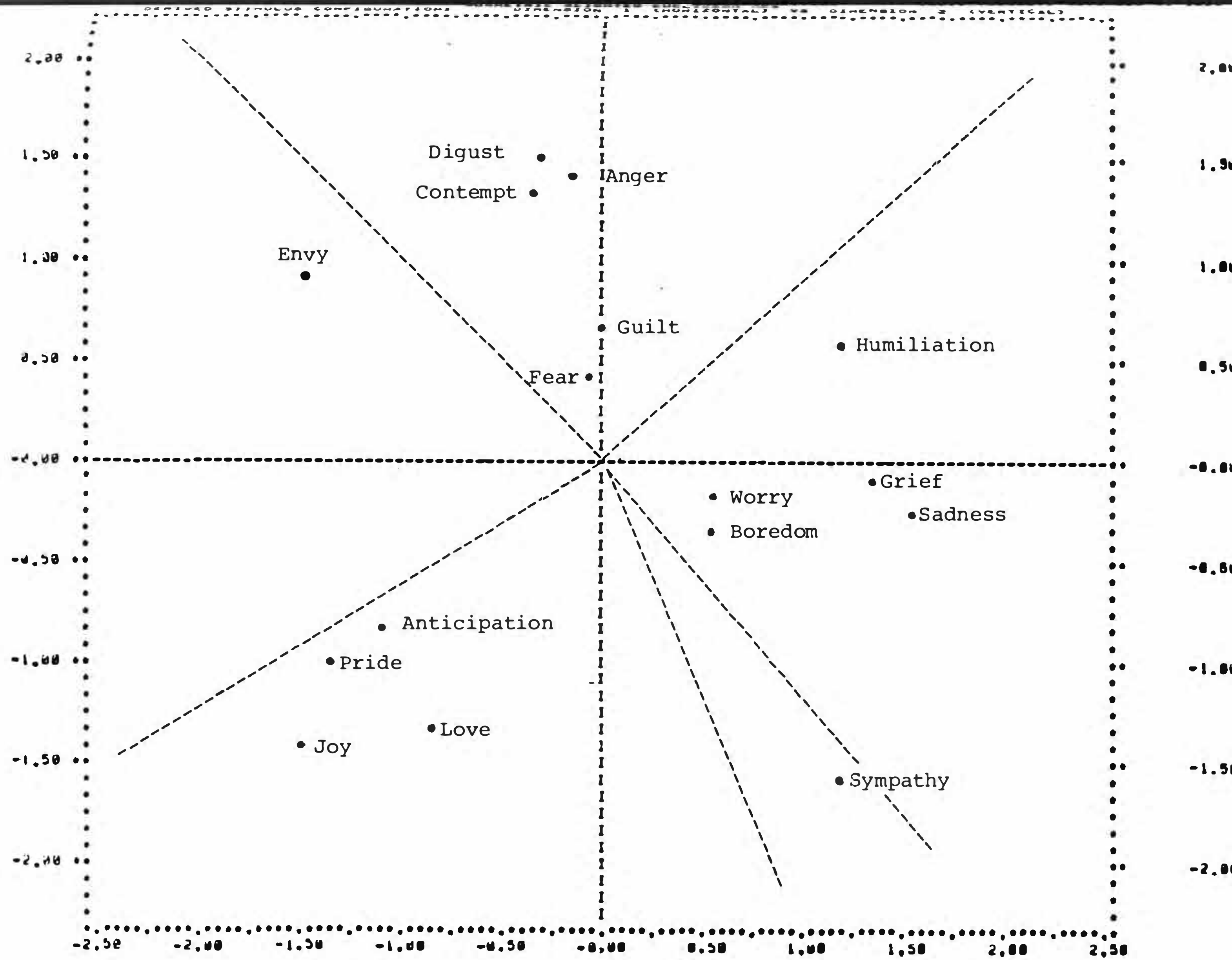


FIGURE 12: 3-D Weighted Ordinal Solution

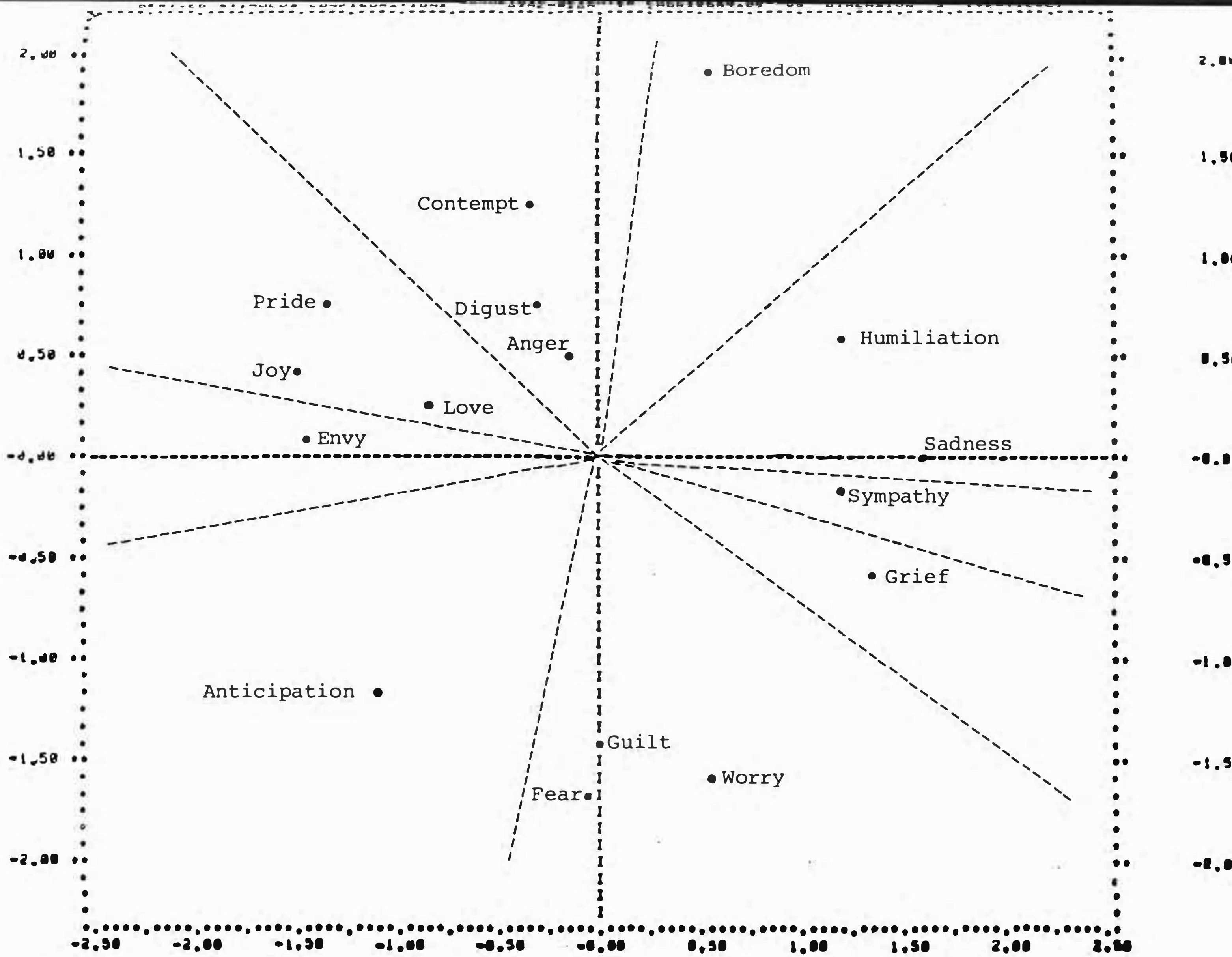


FIGURE 13: 3-D Weighted Ordinal Solution

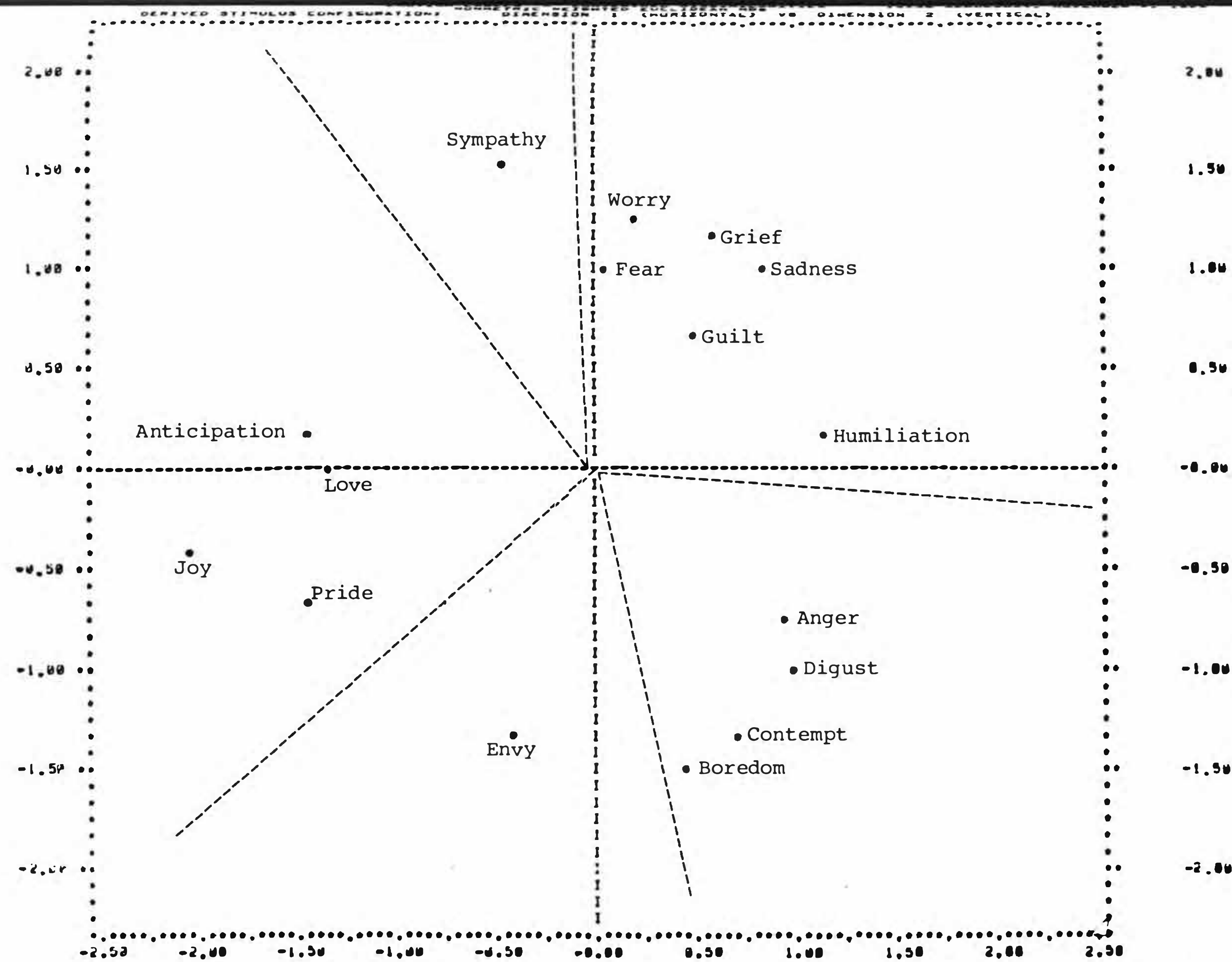


FIGURE 14: 2-D Weighted Ordinal Solution

The following mapping sentence was developed as a plausible interpretation of the derived regional structure within the framework of the conceptual schemes developed for this area.

$$\text{Emotions are} \quad \overset{A}{\left(\begin{array}{l} a_1 : \text{positive} \\ b_2 : \text{negative} \end{array}\right)} \quad \text{affective feelings}$$

$$\text{because of} \quad \overset{B}{\left(\begin{array}{l} b_1 : \text{satisfying} \\ b_2 : \text{dissatisfying} \end{array}\right)} \quad \text{outcomes}$$

$$\text{affecting} \quad \overset{C}{\left(\begin{array}{l} c_1 : \text{self} \\ c_2 : \text{others} \end{array}\right)} \quad \text{and attributed to} \quad \overset{D}{\left(\begin{array}{l} d_1 : \text{internal} \\ d_2 : \text{external} \end{array}\right)}$$

$$\begin{array}{l}\text{factors that} \\ \text{one feels}\end{array} \quad \overset{E}{\left(\begin{array}{l} e_1 : \text{competent} \\ e_2 : \text{incompetent} \end{array}\right)} \quad \text{to handle}$$

This is in fact not a complete mapping sentence as it makes no prediction of a response, but simply serves as the definitional system of the stimulus objects used by subjects to make similarity estimations. That is the similarity of structuples is the direct object of subjective evaluation, rather than the means of predicting similar responses on, for example, a rating scale.

Facet A clearly corresponds to the well-established and obvious dimension of hedonic tone. This however is qualified somewhat in the above system to allow for a distinction between the tone of the feeling itself, and the evaluation of the event precipitating the emotion. This has in fact been given a central role in this scheme with three facets relating to information about the object and cause of the emotion, rather than the emotional state itself. The obtained empirical structure seems to indicate quite clearly that these considerations dominate the more affective aspects considered by most theorists.

Facets B and D relate to the cause of the emotion, that is the outcome which is the reason why one is experiencing the emotional feeling. This outcome is also evaluated positively or negatively (Facet B) and will usually but not always be equivalent to Facet A. For example, Sympathy is a positive feeling regarding a negative outcome, Envy a negative feeling regarding a positive outcome.

The cause of the emotion is attributed to either internal or external factors (Facet D). This uses Rotter's (1966) locus of control terminology to indicate whether the individual does or does not see himself as responsible for his emotional state.

The most important piece of information regarding the object of the emotional feeling is again the distinction between self or others, and this is represented in Facet C. Facet E denotes the competence factor proposed by Davitz (1969), which is the major qualifier of the nature of the emotional state expressed in Facet A. This, together with the control factor in Facet D appears more obviously represented in semantic structure than the relatedness dimension referred to by Davitz.

This 5 Facet structure produces a total of 32 possible structuples, although not all combinations would be expected to correspond to semantic expressions. It would be rather illogical for example for positive feelings for satisfying outcomes to produce a withdrawal response.

The suggested facet structure in fact implies that a complete specification on all five facets will only be relevant for negative emotions. Many positive emotion labels simply refer to general emotional states without any specific cause or object (e.g. joy, cheerfulness, happiness). Others imply that there may be a more specific cause (delight, enjoyment, excitement) but this does not appear a highly salient perceptual distinction. These words have all been classified as positive, satisfying experiences for self which one is confident to handle. While a few positive emotion words do imply different values of some relevance on the other facets (e.g. pride, gratitude, admiration, relief) insufficient of these have been included in the present study to attempt to identify any specific structure of positive emotions.

Table 10 lists all the structuples that are considered feasible, abbreviated with the first letter of each facet, with a dash to indicate any facet that cannot be meaningfully specified. The words used both in this and the Davitz (1969) study are listed as examples in their appropriate classification.

Table 11 summarises the facet structure of the emotion labels selected as stimuli for the present study.

Several points are evident in comparing the formal structure represented in these tables with the spatial configurations shown in Figs 13-15. Firstly in both cases

TABLE 10

Facet Structure of Emotion Words

Structuple	Present Stimulus Set	Davitz (1969)
pss.c	anticipation joy love	affection amusement cheerfulness confidence contentment elation enjoyment excitement friendliness gaiety happiness hope inspiration love passion serenity surprise
pssic	pride	pride
pssec		gratitude
psoec		admiration
psoei		awe reverence
pdoec	sympathy	pity
pds-i		relief
nds--	grief sadness	grief
nds-i	worry	anxiety depression nervousness panic worry
ndsei	fear	fear
ndsii	humiliation	embarrassment
ndse-		frustration
ndsec	anger	anger dislike hate impatience irritation resentment
ndoec	contempt disgust	contempt disgust
ndoii	guilt	guilt remorse shame
nso-i	envy	jealousy

TABLE 11

Comparative Facet Structure of Selected Stimulus Set

	A	B	C	D	E
Contempt	n	d	o	e	c
Disgust	n	d	o	e	c
Humiliation	n	d	s	i	i
Anger	n	d	s	e	c
Envy	n	s	o	e	i
Love	p	s	s	-	c
Fear	n	d	s	e	i
Joy	p	s	s	-	c
Pride	p	s	s	i	c
Worry	n	d	s	-	i
Anticipation	p	s	s	-	-
Guilt	n	d	o	i	i
Boredom	n	d	s	-	-
Sadness	n	d	s	-	-
Sympathy	p	d	o	-	c
Grief	n	d	s	-	-

KEY: (FOR TABLES 10 & 11)

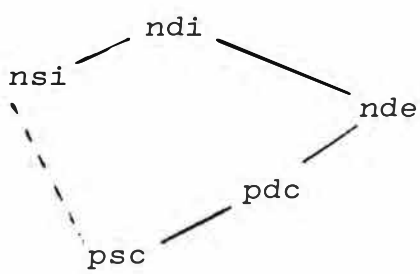
FACET A p = positive
n = negative
FACET B s = satisfying
d = dissatisfying
FACET C s = self
o = other
FACET D i = internal
e = external
FACET E c = competent
i = incompetent

negative emotions are more specifically distinguished than positive emotions. The reasons for expecting this were discussed previously.

Secondly, some facets clearly play a more dominant role than others. The major regional clusters evident in the MDS configuration appear to correspond to combinations of Facets A, B and E. This is in fact clearer in the 2 - D solution than the 3 - D solution. That is, these three clearly seem to represent 'semantic principal components' of emotion labelling. To generate a circumplex from a three-facet structure requires specification of two facets to be the more dominant factors. Facet A is one obvious choice, this being clearly the most important aspect distinguishing emotion labels. Facet B seems a logical second choice, largely because of its strong dependence on Facet A. This produces the following expected order among the eight primary structuples

psc
psi
pdi
pdc
ndc
ndi
nsi
nsc
psc

The stimuli selected for this study contain examples of five of these structuples only. The cyclic order among these five structuples, listed below, is a subset of the full cyclic order listed above



It is rather difficult to identify a dominant ordering of the regions in the 3-D solution, as the additional dimension enables a more complex juxtaposition of regions. If the present interpretation is correct the dominant conceptual features of the data structure can be represented in two dimensions, and the third dimension is only necessary to represent some distortions in this structure. The obtained 2-D solution does not correspond perfectly with the predicted structure, as the regions nsi and pdc have been interchanged. These are the single-word regions 'envy' and 'sympathy', and especially as the fit in this dimensionality was not unduly good, this does not appear to represent a major departure from expected structure. The perfect location of words within the five clearly distinct regions provides quite a strong reassurance that for at least three of the defined facets there is good empirical confirmation of the predicted theoretical structure.

Confirmatory Analyses of Facet Structure

This could thus be regarded as the conclusion of the exploratory phase of the analysis. A plausible theoretical structure has been developed which appears reasonably consistent with the obtained empirical structure. The next phase involves some attempts at a more explicit quantitative confirmation of the structural hypotheses implied by this theoretical analysis.

The first analysis attempted was a direct comparison of the role of different facets and facet combinations in the assessment of similarity. This involved partitioning the set of stimulus pairs into groups with equivalent values on some facets and contrasting values on other facets. The significance of the contrast facet (or facets) was assessed by a Mann-Whitney U test of the difference between the rank ordered similarity values for stimuli pairs equal or different on this facet.

The set of contrasts examined is detailed in Tables 12 and 13. Table 12 lists the mean similarity for 'same' and 'different' values of the contrast facet respectively, for stimulus pairs having the same values on the common facets. Table 13 shows the results of Mann-Whitney U tests of the significance of these comparisions. The U values have been transformed to standard normal deviates for ease of comparison across the table. Significant values ($p < .05$) are underlined. The z-transformation is not very accurate for some of the low sample sizes occuring towards the end of the table. Exact probabilities have been used as the significance criteria in these cases, even though the z-scores are the values listed.

Table 13 shows the dominant effect exerted by Facet A. Facets A and B are the only ones to achieve significance as single facets for the majority of subjects, Facet B no doubt due to its large overlap with Facet A. There are only two subjects without significant Facet A effects. S5 shows a trend approaching significance (it would be significant one-tailed).

TABLE 12

Mean Similarity by Facet Structure

Common Comparison						A				AB			AE			ABE	
	A	B	C	D	E	B	C	D	E	C	D	E	B	C	D	C	D
1	3.29 2.00	3.15 2.16	2.64 2.76	3.23 2.67	3.00 2.22	3.35 3.07	3.42 3.15	3.54 3.30	3.68 3.07	3.43 3.24	3.43 3.38	4.33 2.75	4.33 2.57	4.25 3.27	3.80 3.50	4.43 4.20	4.25 4.00
2	5.62 3.24	5.29 3.62	4.38 4.69	7.15 6.60	3.36 3.12	5.62 5.57	5.91 5.31	7.09 6.30	6.47 5.80	5.83 5.33	6.86 6.88	7.50 5.33	7.50 4.71	7.38 5.82	7.20 6.50	7.71 7.20	7.25 9.00
3	4.95 3.51	5.22 3.20	4.34 4.23	4.31 4.67	4.22 4.03	5.20 4.07	5.42 4.67	4.64 5.50	4.95 4.33	5.50 4.76	4.29 5.63	5.83 4.25	5.83 3.43	5.75 4.36	5.60 5.25	5.86 5.80	5.25 5.50
4	5.77 6.20	5.98 5.95	6.00 5.93	5.92 6.73	6.09 5.86	5.94 5.14	5.94 5.59	5.45 6.50	5.31 5.60	5.77 6.19	5.14 6.38	5.58 5.17	5.58 4.86	5.37 5.27	6.00 5.50	5.14 6.20	7.00 4.00
5	4.26 3.60	4.18 3.69	4.01 3.89	4.85 3.53	4.26 3.86	4.33 4.00	4.48 4.03	4.73 3.90	5.42 3.60	4.57 4.00	5.00 4.00	6.00 3.50	6.00 4.23	5.87 5.09	5.60 4.25	6.29 5.60	5.75 5.00
6	6.15 4.38	5.66 4.96	5.32 5.36	5.77 6.60	5.97 5.46	6.02 6.64	6.52 5.78	6.00 7.00	7.00 6.53	6.43 5.43	5.29 7.00	7.00 7.67	7.00 7.00	7.50 6.64	5.20 7.25	7.57 6.20	5.00 7.50
7	3.51 1.64	3.40 1.76	2.77 2.51	4.46 3.00	2.77 2.40	3.76 2.57	3.91 3.09	4.55 3.30	4.11 3.47	3.97 3.48	5.43 3.25	5.00 3.58	5.00 2.57	4.88 3.55	5.20 3.25	52.9 4.60	5.75 3.00
8	6.31 3.80	6.80 3.22	4.69 5.71	5.61 6.06	5.58 5.26	6.88 4.21	6.21 6.41	6.45 7.10	7.11 6.47	6.47 7.48	8.14 7.38	8.42 7.42	8.42 4.86	8.00 6.45	7.80 7.00	8.43 8.40	8.25 8.00
9	6.46 4.09	6.37 4.20	5.23 5.54	6.61 6.73	6.32 4.77	6.75 5.43	7.00 5.91	7.27 6.80	7.89 6.13	7.13 6.19	8.14 6.75	8.58 6.42	8.58 6.72	8.00 7.82	8.80 7.25	8.29 9.00	9.00 7.50
10	4.98 3.31	5.14 3.13	4.57 3.80	4.61 4.93	4.54 4.09	5.45 3.29	5.67 4.28	5.09 6.00	5.79 4.93	5.90 4.81	5.57 6.87	7.42 5.42	7.42 3.00	7.38 4.64	7.00 4.75	7.71 7.00	6.75 7.00
11	5.78 3.93	5.69 4.03	4.78 5.11	5.77 6.00	5.32 4.94	5.86 5.50	6.30 5.25	6.36 6.60	6.89 5.53	6.20 5.38	6.00 6.37	7.92 4.92	7.92 5.14	7.88 6.18	7.00 7.00	8.00 7.80	7.75 6.50

TABLE 13

Rank Order Comparisons of Facet Structure
(Standard Normal Transformations of Mann-Whitney U Scores)

Common						A				AB			AE			AB	AE	ABE	ABE
Comparison	A	B	C	D	E	B	C	D	E	C	D	E	B	C	D	CD	CD	D	C
1	5.23	3.92	-0.72	0.83	2.00	0.79	0.83	0.18	1.36	0.46	-0.06	2.93	2.79	1.62	0.13	0.05	0.44	1.00	0.26
2	4.19	3.12	-0.62	0.71	1.08	0.36	0.86	0.58	1.05	0.60	0.49	2.01	2.03	1.05	0.27	1.93	0.44	0.98	0.18
3	3.23	4.25	0.29	-0.25	0.32	1.47	1.66	-0.93	0.65	1.12	-1.00	1.40	1.92	1.21	0.12	0.15	0.67	0.00	0.33
4	-0.82	0.10	0.34	-0.67	0.50	0.98	0.50	-0.79	-0.44	-0.48	-0.83	0.06	0.70	0.38	0.63	-1.51	0.88	0.50	-0.26
5	1.75	1.25	0.05	2.09	0.59	0.58	0.64	1.30	2.66	0.72	1.26	3.09	2.02	0.67	1.68	-0.59	-1.00	1.97	0.43
6	4.05	1.78	-0.01	1.16	0.97	-0.39	1.47	-1.27	0.38	1.53	-1.74	0.92	0.58	1.58	-2.19	0.15	-0.22	1.47	1.50
7	5.42	4.71	0.60	2.51	0.35	2.27	1.83	1.94	0.93	0.96	2.79	1.90	2.41	1.30	1.89	1.16	0.44	1.75	0.75
8	4.45	6.40	-1.87	-0.07	0.76	3.34	-0.45	-0.14	1.24	1.39	1.09	2.15	2.87	1.33	0.88	1.36	0.44	1.97	0.46
9	4.11	4.23	-0.28	0.35	2.71	2.20	1.89	0.80	3.15	1.17	1.76	3.65	3.05	0.17	1.76	1.28	0.56	2.19	-1.60
10	3.03	3.66	1.43	-0.50	1.12	2.57	2.07	-1.30	1.44	1.44	-1.56	2.81	3.11	1.97	1.16	0.18	0.56	1.67	1.07
11	3.76	3.41	0.52	-0.14	0.75	0.56	1.81	-0.07	1.90	1.19	-0.47	3.32	2.70	1.40	0.50	0.44	0.44	0.98	0.00
(n_1, n_2)	(65,55)	(65,55)	(65,55)	(13,15)	(31,35)	(51,14)	(33,32)	(11,10)	(19,15)	(30,21)	(7,8)	(12,12)	(12,7)	(8,11)	(5,4)	(2,49)	(1,18)	(4,2)	(7,5)

NOTE: Underline values refer to differences significant at $p < .05$.

S4 on the other hand appears to show no effects of this or any other facet combination. If this subject is responding rationally it is clearly not using a system at all similar to those considered here. While there are one or two scattered significant effects for Facets D and E a much more consistent picture emerges if we select Facet A as a common facet. (i.e. we compare only positive emotions with positive emotions and negative emotions with negative emotions). Facet B shows trends in a consistent direction for all but one subject, reaching significance for four subjects (S7, S8, S9, S10). This is in this case testing for 'sympathy' to be seen as distinct from other positive emotions, and 'Envy' as distinct from other negative emotions. Facet C shows a fairly consistent although clearly small effect, with only one significant comparison. Facet E likewise shows a consistent but small effect.

When both Facets A and B are selected as common facets the effect of Facet E is much more apparent. All subjects apart from S4 show recognizable trends, with seven of them reaching significance. Clearly then the facets operate in a sequential fashion with Facet E only being utilised for stimuli equated on Facets A and B. Strategies involving sequential consideration of effects, with the possibility of a decision being made at any point in the sequence, are defined as <u>Lexicographic</u> decision rules. The evidence from studies of choice behaviour suggests that such strategies may be quite common, and of course represent a critical violation of the independent dimensions assumption of MDS models.

Facet C does not show any stronger effects as one increases the common facet structure. While it does seem to show up as contributing a small amount it appears to operate as an independent rather than lexicographic effect. The evidence for Facet D is even more doubtful. There is some evidence for it being involved as a contrast when all three major facets are in common, but with such low sample sizes one cannot place too much emphasis on this.

Constrained MDS Analysis

Having determined that at least some facet combinations from the hypothesized facet structure are significant predictors of similarity for at least some subjects it now remains to be shown whether the overall facet structure is a reasonable basis for interpretation of the spatial structure derived from the similarities. The technique of constrained MDS will provide a qualitative empirical evaluation of the degree to which it is possible to derive a spatial representation of the data that is consistent with this theoretical interpretation. A quantitative measure of the appropriateness of this theoretical interpretation can be obtained by comparing the goodness of fit of a spatial configuration constrained to conform to the structural hypothesis derived from the theory with the goodness of fit obtained under unrestricted scaling conditions.

Most restricted MDS programs express their constraints in terms of the projections of points on the dimensions of the configuration, for example constraining them to be in a specific order, or to have a prespecified constant value.

Bentler and Weeks (1978), Bloxam (1978) and Noma and Johnson (1977a,b) describe procedures of this type. Lingoes (1979) however has criticized this approach as being limited by the lack of dimensional uniqueness.

Facet theory on the other hand leads to nondimensional structural hypotheses which apply constraints directly on the interpoint distances, while ensuring that a spatial representation will still be possible. Lingoes and Borg (1978) describe a program CMDA (Constrained/or Confirmatory Monotone Distance Analysis) which enables constraints of this sort to be placed on the order of distances in the configuration. The constraints are expressed in terms of one or more theoretical proximity matrices (more than one is required if the constrained values are not matrix conditional) and the scaling algorithm simultaneously operates on both these and the obtained proximity data. This is done by partitioning the loss function into a component that measures the satisfaction of the monotonicity requirements (i.e. from the data) and one or more components that are related to the imposed constraints. The object is to satisfy the constraints component(s) to the maximum extent possible and to derive the best compromise solution for the monotonicity requirements subject to this. This is achieved by applying a variable penalty function on the components, which starts by giving equal weighting to both data and constraints, but with each iteration gives gradually more weight to the constraints until finally one is scaling these alone. (This is quite a different strategy to the Noma and Johnson procedure which alternates between the two error criteria). When this loss function has been minimized to some acceptable criterion the goodness of fit of the proximity matrix to this theoretically consistent configuration can be evaluated using standard procedures (e.g. either Kruskal's stress or Guttman's coefficient of alienation).

Lingoes and Borg (1980) describe a basis by which one can test statistically whether this represents a significant loss of variance explained from the unrestricted solution, and thus whether the theory can be maintained as a plausible explanation of the spatial configuration. The unrestricted solution is used as the starting configuration in the procedure both to minimise problems of local minima (in terms of satisfying data constraints) and to serve as an appropriate comparison configuration for the statistical evaluation of the final configuration.

Results

The CMDA program was run on the Hewlett-Packard 3000 computer at the Gippsland Institute of Advanced Education. Subjects were analysed individually, although since it is a common configuration that is being evaluated, the common configuration obtained from the ALSCAL scaling of all 11 subjects was used as the initial configuration in all cases. Scalings were performed in three and two dimensions, as it appeared that even though three dimensions were required on goodness of fit considerations, much of the theoretical information was retained in the two-dimensional solution. Two levels of constraints were applied, firstly using a theoretical structure derived from just the three most significant facets, then from the total five facet structure. The theoretical proximity matrices for these two conditions are shown in Tables 14 and 15. These matrices incorporate the assumptions that differences on the major facets A, B & E will dominate differences on the minor facets C & D, and that missing facets can be considered as lying on the neutral point between the two possible values on that facet. Thus in Table 15, values 9,8,7 and 6 all refer to differences less than that of a contrasting value on one major facet. Values 7 and 6 refer to cases including a missing value on a major facet, or a difference of one half of a major facet. The major and minor facets are assumed to operate lexicographically, and thus minor facet differences do not produce a cumulative effect greater than the next major facet difference (or half difference).

Table 16 shows the various goodnesses of fit criteria for each of the four sets of scaling conditions. The first two columns show the stress values (Kruskal's Stress Formula (1) is used) measuring the monotonic fit of the two proximity matrices to distances in the final configuration; S(U, P) and S(C, R1) relating to proximity and constraints pseudodata matrices respectively.

The next three columns contain product moment correlations between data P and rank-images of the distances in both initial and final configurations P * (IC), P * (FC) and between these two sets of rank-images. The statistical test derived by Lingoes and Borg (1980) tests the null hypothesis that R(P, P* (IC) = R(P, P* (FC). If this is not disconfirmed at

TABLE 14

Three Facet Constraint Matrix

Contempt															
Disgust	6														
Humiliation	4	4													
Anger	6	6	4												
Envy	4	4	4	2											
Love	2	2	0	2	2										
Fear	4	4	6	4	2	0									
Joy	2	2	0	2	2	6	0								
Pride	2	2	0	2	2	6	0	6							
Worry	4	4	6	4	4	0	6	0	0						
Anticipation	1	1	1	1	3	5	1	5	5	1					
Guilt	4	4	6	4	4	0	6	0	0	6	1				
Boredom	5	5	5	5	3	1	5	1	1	5	2	5			
Sadness	5	5	5	5	3	1	5	1	1	5	2	5	6		
Sympathy	2	2	2	4	0	4	2	4	4	2	3	2	3	3	
Grief	5	5	5	5	3	1	5	1	1	5	2	5	6	6	3

KEY

Differ on no facets 6

Differ on missing facet 5

One facet 4

One and missing facet 3

Two facets 2

Two and missing facet 1

Three facets 0

TABLE 15

Five Facet Constraint Matrix

Contempt															
Disgust	9														
Humiliation	4	4													
Anger	8	8	4												
Envy	2	2	4	4											
Love	1	1	0	1	1										
Fear	4	4	8	5	4	0									
Joy	1	1	0	1	1	9	0								
Pride	1	1	0	1	1	8	0	8							
Worry	4	4	8	4	4	0	8	0	0						
Anticipation	0	0	1	1	3	4	1	7	6	1					
Guilt	4	4	8	4	4	0	8	0	0	8	1				
Boredom	6	6	7	6	3	1	6	1	1	7	1	6			
Sadness	6	6	7	6	3	1	6	1	1	7	1	6	9		
Sympathy	4	4	1	4	0	4	1	4	4	1	3	2	1	3	
Grief	6	6	6	6	3	1	6	1	1	7	1	6	9	9	3

KEY

Differ on no facets 9
Differ on minor facets only 8
Major with missing facet 7
Major with missing facet, and minor facets 6
Differ on one major facet 5
One major and minor facets 4
One major, major and missing, and minor facets 3
Two major facets 2
Two major and minor facets 1
Three major facets 0

TABLE 16

CMDA Fit Coefficients

(A) Three Facet Constraints - 3-Dimensional Solution

	S(U,P)	S(C,R_1)	R(P,P*(IC))	R(P,P*(FC))	R(P*IC)(P*FC)	T
1	.262	.004	.540	.520	.552	.288
2	.285	.005	.573	.463	.597	1.639
3	.328	.005	.500	.416	.640	1.249
4	.320	.006	.275	.135	.547	1.656
5	.343	.005	.409	.247	.602	2.165
6	.325	.005	.408	.365	.594	.574
7	.264	.004	.580	.557	.582	.361
8	.281	.003	.534	.523	.614	.159
9	.269	.003	.530	.503	.554	.380
10	.274	.003	.554	.412	.565	1.998
11	.286	.005	.730	.443	.591	5.027

(B) Three Facet Constraints - 2-Dimensional Solution

1	.402	.010	.450	.496	.533	.607
2	.450	.013	.532	.384	.515	1.944
3	.474	.014	.495	.349	.566	1.959
4	.484	.013	.137	.011	.539	1.433
5	.512	.016	.422	.147	.529	3.397
6	.477	.016	.394	.273	.529	1.466
7	.393	.010	.564	.502	.553	.891
8	.429	.010	.673	.504	.580	2.760
9	.422	.015	.476	.469	.543	.097
10	.434	.015	.591	.340	.526	3.451
11	.456	.015	.720	.417	.572	5.117

TABLE 16 (Cont'd)

CMDA Fit Coefficients

(C) Five Facet Constraints - 3-Dimensional Solution

	S(U,P)	S(C,R_1)	R(P,P*(IC))	R(P,P*(FC))	R(P*(IC)(P*FC)	T
1	.265	.007	.540	.518	.512	.298
2	.292	.006	.572	.473	.595	1.488
3	.327	.007	.503	.422	.568	1.112
4	.322	.007	.301	.023	.514	3.234
5	.349	.008	.415	.230	.550	2.318
6	.322	.006	.399	.337	.557	.782
7	.273	.007	.593	.553	.582	.626
8	.286	.009	.543	.556	.570	.194
9	.278	.008	.548	.514	.593	.507
10	.286	.008	.570	.417	.558	2.177
11	.296	.006	.730	.429	.553	5.047

(D) Five Facet Constraints - 2-Dimensional Solution

1	.405	.015	.450	.427	.504	.298
2	.435	.018	.532	.417	.548	1.566
3	.414	.013	.495	.432	.578	.877
4	.477	.019	.137	-.031	.506	1.857
5	.502	.020	.422	.218	.586	2.679
6	.478	.018	.394	.323	.568	.896
7	.372	.010	.564	.521	.586	.650
8	.425	.015	.673	.519	.589	2.555
9	.433	.017	.476	.470	.547	.078
10	.430	.016	.591	.362	.558	3.268
11	.451	.019	.720	.397	.582	5.523

the chosen significance level then it is appropriate to regard the theoretical (final) structure as a plausible explanation of the initial structure.

The stress levels for the pseudodata matrices show that the constraints can be almost perfectly satisfied in both 3 and 2 dimensions. The data stress values are rather high, averaging around .3 for 3 dimensions, .4 for 2 dimensions. While these would perhaps be considered unacceptably high applying Kruskal's (1964) guidelines, one cannot apply these constantly across all scaling conditions. As constrained scaling solutions they are not local minimum solutions for the data, and other factors such as number of points and dimensions also effect stress values. In addition, as Noma (1978) demonstrates, the notion of acceptable stress levels of a solution should be dependent on some measure of the level of inherent noise in the data. He demonstrates that the intersession response variability is a suitable basis for assessing this, and describes the results of a Monte Carlo study relating this to stress values, using artificial dissimilarity data constructed with varying degrees of error. While this measure is not available in the present study it seems reasonable to expect a fairly high level of response variability in such a cognitively complex and ambiguous empirical domain.

The three facet constraints have resulted in relatively minor increases in stress levels for most subjects in the 3-dimensional solution, with only three subjects, S5, S10 and S11, showing significant reductions ($T < 1.998$, df = 117, $p < .045$). S5 was identified in the previous analysis as showing limited use of the primary evaluative facet only. S10 and S11 did display significant effects of these three facets on similarity, and their lack of fit here most plausibly indicates some nonindependent (lexicographic?) strategy for the combination of facets.

The situation is significantly worsened in 2 dimensions, with six subjects showing substantial reductions in fit ($T < 1.94$, df = 117, $p < .051$). This is however a fairly degenerate theoretical solution for the present stimulus set as only five clusters of stimuli are distinguished by these three facets, and thus, under the primary approach to ties adopted in confirmatory scaling only a small number of levels of interpoint distances are allowed. The fact that three subjects S1, S7, S9, showed good fits even in 2 dimensions provides some support for the three major facets as constituting at least part of the structure for most subjects.

When the full five facet structure is used to generate constraints the results do show a lesser stress increase for most subjects, indicating that the extra facets do make some additional contribution to the prediction of configurational distance. The same three subjects show significant increases in the 3-dimensional solution as in the three facet case, with the addition of the poorly fitting S4.

This improvement is even more marked in the 2-dimensional solution. This shows quite a clear separation into two clusters of subjects that can or can not be scaled by the full model in two dimensions. Five subjects (S1, S3, S6, S7, S9) show minimal stress differences, with a further two showing larger but nonsignificant increases (S2, S4). The remaining four subjects show quite substantial increases in stress when constrained by the theoretical solution. Only one subject, S8, appears to have shown any significant drop in fit to the theoretical configuration in the change from three to two dimensions.

Discussion

Definite conclusions on the degree of confirmation of the predicted theoretical structure are limited by the high degree of residual noise in the data, and the incomplete representation of facet combinations in the stimulus set. For an individual subjects analysis this means that capitilization on chance effects are difficult to distinguish from variance consistent with the predicted model. However the confirmatory scaling analysis has suggested that to the extent the semantic structure of emotion labelling can be represented as a common spatial configuration for this group of subjects, the theoretical facet analysis derived here constitutes a plausible explanation for this structure.

There is some evidence that for several subjects who show substantial agreement with the content of the predicted facets, the MDS model is far from an adequate representation of the way these are combined.

The results of this study however appear sufficiently encouraging to suggest that a more comprehensive replication would be worthwhile. More ideal conditions to evaluate this and other possible theoretical structures of emotion labelling would include the use of a larger stimulus set, giving a more representative coverage of the theoretical groupings suggested here. This in itself would provide a more reliable basis for estimation of each subject (if complete presentations are used), and this could be usefully increased further by the use of one or more replications.

CHAPTER VII

CONCLUSION

This thesis has considered a variety of approaches to the utilization of MDS procedures in evaluating and contributing to existing substantive psychological theory. A major position of this thesis has been that such questions cannot be meaningfully examined without considering them in the light of wider metatheoretical issues involved in quantification in psychology in general. The major positions on these metatheoretical issues were reviewed in Part I.

The major question concerns the formal statement of the requirements for the justification of a psychological measurement procedure. It has been argued here that this can only come from the measurement theoretical approach as represented by, for example, Krantz, Luce, Suppes and Tversky (1971). This implies a need for the scale values to be clearly related to basic empirical observations via a set of justifiable axioms, and formal proofs that such a representation can be derived from these axioms and of the uniqueness of any scale so derived.

One can also identify some variations within this approach over the nature of data that can be used as basic empirical operations, particularly regarding the status of direct subjective estimates of numerical magnitudes. These have been variously regarded as indirectly or directly related to some theoretical level of 'sensation magnitude', or as the direct behavioural data we as psychologists seek to explain. The latter two positions imply that this dat a should be explained at least in terms of its interval level properties, while the first would imply that weaker properties, such as weak ordering or continuity, should be used. This is the position which is regarded here as theoretically most defensible, although it may also be considered somewhat conservative, in the light of evidence that magnitude estimates do display sufficient cross-situational consistency to be used in interval level predictions. The choice then is as much an empirical as a theoretical one, depending on the richness and generality of models that can be generated by each approach.

By way of a summary and conclusion, it is proposed to briefly review the various positions taken on the specific issue of the role of MDS in psychological research by the major contributors to theory and methodology in this area in terms of their positions, usually implied rather than stated, on the broader issues considered above.

One should first, perhaps, refer to a distinction, expressed in various forms by Gregson (1975, 1976) and Krantz and Tversky (1975), between modelling the perception of stimulus attributes and using MDS as a model of cognitive processes. The former approach assumes that natural (i.e. physically obvious and distinct) attributes can be regarded as basic psychological features that enter in a perceptual strategy, and involves the formulation, analysis and testing of rules for combining these features. The set of possible combination rules is thus not restricted to those that obey the requirements of a distance metric as would be necessary for a MDS. The second approach in contrast makes the assumption that the ADM model is a plausible model of the combination of unknown features and uses MDS to attempt to extract dimensions that correspond to these features.

This distinction was at one stage referred to as a content-distance model dichotomy (Ekman & Sjoberg, 1965), because many similarity models start from a heuristic definition of similarity as the quotient of common stimulus content divided by the total content. However, as Gregson (1976) comments this distinction is a somewhat unsatisfactory basis for classifying similarity models as it obscures some more fundamental differences within these categories. Some content models for example can be written as normalised distances and thus share some of the formal structure of distance metrics, although they do not in fact have distance properties. Some evidence from comparative studies of various types of similarity models suggests that content models which can be written as normalised distances are the most empirically successful similarity models, while content models which can not be transformed to this form and distance model, show a fairly equivalent but lesser degree of fit (Gregson 1975, 1976, Eisler and Roskam 1977a,b).

Another means of distinguishing between similarity models is based not on their mathematical structure but the way the primitives of the model are related to physical attributes. At least three basic types of representations can be distinguished here. The most common is the dimensional representation where stimuli are represented as values on a set of dimensions. This is the form that must be assumed by distance models, and may be used by some content models. More usually however content models are expressed in set theoretic form based on measures of sets of features possessed in common or disjointly by two stimuli. (Gregson, 1975).

A different type of set theoretic representation for sets of stimulus features was presented by Tversky (1977). Tversky proposed that the similarity function should be predicated not just on total and common content, but on all three sets A U B, A-B, B-A, to allow for assymetric simmilarities. Tversky quoted several empirical examples where one can encourage subjects to generate assymetric similarities, although there is no firm evidence to suggest that this is a universal feature. Tversky's formulation implies that features are measured on a presence/absence basis, as otherwise the two disjoint sets would cancel out.

Our choice between the various competing similarity models is thus dictated not simply on the basis of the goodness of fit of their underlying mathematical structure, but also the appropriateness of their representations to specific data sets.

This is an important consideration when one also considers that direct similarity modelling studies must of necessity use stimuli whose physical dimensions are perceptually obvious and distinct. It is quite likely that the nature of the variation constructed into the stimulus set could determine which of the representations described above is most appropriate. Other factors limiting the applicability of these studies are their dependence on the validity of the interval level information in estimates of similarity magnitudes, and the fact that many studies have used data pooled over subjects.

This similarity modelling approach has not been considered to any great extent in this thesis. While the results do have some disquieting implications for MDS, the considerations above mean that these are insufficient to invalidate their use.

The second major strategy referred to above recognizes the advantages in scaling stimuli whose basic features are unknown if the necessary assumptions of the scaling model can be accepted as reasonable. Shepard (1974) referred to the discovery of hidden structure as the most important objective of the use of MDS in analyzing similarity data. One can distinguish a variety of different attitudes regarding the conditions which need to be satisfied before MDS can be used for this purpose.

The position taken by Shepard and Kruskal when initially proposing the nonmetric scaling model was that the monotonic transformation avoided the need to impose any conditions at all. Shepard argued that since one could consider both similarity and distances as special cases of the more general notion of proximities, then order relations on one should map into order relations on the other. Thus one needs only rely on the arguments of Abelson and Tukey (1959), which showed that a sufficient number of ordinal contraints could generate metric information, to justify using the distance configuration that best reproduces the similarity ordering as a representation of the metric structure of the similarities.

However, while nonmetric scaling programs can successfully retrieve fairly stable metric representations, so much freedom is allowed in the form of the arbitrary monotonic transformations that systematic departures from the scaling model have little chance of being observed.

The most rigorous conditions for justifying MDS are of course those that have been identified by the measurement-theoretic analysis of the ADM model, and thus the most defensible procedure would be to demonstrate that these can be satisfied for the given data set in every scaling application.

The basic underlying model of MDS, the additive difference model, states two basic properties, intradimensional subtractivity and interdimensional additivity, that must necessarily be satisfied by any psychological dimensions if they are to enter into a cognitive judgement that can be considered as having distance properties. One could in fact state that an even more basic requirement is that stimuli be perceived and evaluated in a dimensionally organised fashion. That is, we can state an even more general requirement than subtractivity denoted decomposability, as

$$d(x,y) = F\ [\phi_1(x_1,y_1),\ldots\ldots\ldots\ \phi_n(x_n,\ y_n)]$$

which states that the distance between points must be a function of component wise contributions. Equation 2 then specifies that these componentwise contributions be the absolute value of the scale differences. The major result of the studies which have tested the predictions of the ADM has been the systematic rejection of even this more general requirement of decomposability (Tversky and Krantz, 1970; Krantz and Tversky, 1975). The general finding was instead the identification of alternative types of combination rules based on interactions between dimensions.

While this result appears to critically weaken the case for using MDS as a cognitive model, there are a number of qualifications that should be attached to this conclusion.

Firstly these studies must again of necessity involve only classes of stimuli with highly distinct dimensions. In these cases, where we have sufficient knowledge about dimensions to evaluate the scaling model, the evidence does suggest it is often inappropriate, and, as Experiment 1 has demonstrated, this can lead to plausible but misleading solutions. However there is also much evidence to suggest that these cases are quiteatypical of stimuli where we have a more genuine need for MDS, i.e. where the perceptual features are not easily distinguished. Thus it may well be that it is precisely the cases we use to evaluate the MDS model that show the greatest systematic departures from its assumptions.

Another key feature in interpeting the significance of these results is the mea ning that is attributed to the dimensions of the scaling model. If they are to be matched in a one to one basis with the features by which stimuli are perceived and structured as is the case in the treatments of Tversky and Krantz, and as is also implied in most applications of MDS procedures, then the negative conclusions discussed above are clearly of critical significance.

However an alternative view is represented by the approach of Lingoes, Guttman and their associates, where the dimensions of a MDS configuration are regarded as arbitrary reference axes, generated for computational convenience only, and not necessarily to be accorded any psychological interpretation. Instead other formal aspects of the configuration such as directions, regions and manifolds may also be used as the basis for any psychological interpretation. Guttman's facet theory approach shows how a similarity model based on a (usually larger) set of qualitative features could be embedded in this underlying dimensional structure.

The use of confirmatory analysis principles in conjunction with nondimensional hypotheses shows some of the spirit if not the letter of measurement theory consistent approaches to the application of MDS to investigate empirical structure, and this appears the currently most promising line of development in this area. However some greater knowledge of the conditions under which one can justify the basic assumptions of a parallel between the fit of structural hypotheses and the appropriateness of corresponding psychological theories is clearly necessary before one can give this more unqualified approval.

It should be clear from the results reviewed in this thesis that little support can be given for what might be termed a methods-oriented approach to MDS, which characterises the contributions of several prominent researchers in this field, such as Shepard, Young, Carroll and Kruskal.

This approach accepts the view that similarity can be directly equated with distance in a derived configuration, either directly as in metric MDS models such as INDSCAL, or indirectly via the retrieval of metric from nonmetric information in nonmetric MDS, and thus that psychological properties of the representation can be inferred from statistical properties of the solution. The 'perceptual structure' interpretation attributed to the dimensions of INDSCAL is a prime example of the unrealistic attitude to theoretical explanation this perspective can produce.

The methodological development within this perspective has followed a trend for more and more flexible scaling models, with sufficient generality to cope with any experimental situation, but with little attempt to demonstrate that a given extension can be justified. Sophisticated statistical procedures using regression or discriminant analysis techniques to aid dimensional interpretation are of little value if no attempt has been made at a prior specification of expected structure. While many very efficient procedures have been developed with many appealing statistical properties, improvements of this sort alone cannot guarantee meaningful results. They may be of some value used in a more limited exploratory mode, but most studies applying MDS to substantive areas seem unwilling to be this cautious in their interpretations (see Forgas and Menyhart (1979) for a recent example which even treats the perceptual structure hypotheses as an established fact).

REFERENCES

Abelson, R.P., & Tukey, J.W., Efficient conversion of non-metric information into metric information. *Proceedings American Statistical Association, Social Statistics Section,* 1959, 226-230.

Alderfer, C.P., An empirical test of a new theory of human needs. *Organizational Behaviour and Human Performance,* 1969, 4, 142-175.

Anderson, N.H., Functional measurement and the psychophysical judgement. *Psychological Review,* 1970, 77, 153-170.

Arabie, P., Concerning Monte Carlo evaluations of non-metric MDS algorithms. *Psychometrika,* 1973, 38, 607-608.

Arabie, P., Random versus rational strategies for initial configurations in nonmetric multidimensional scaling. *Psychometrika,* 1978, 43, 111-113, (a).

Arabie, P., The difference between "several" and "single". A reply to Spence and Young. *Psychometrika,* 1978, 43, 119, (b).

Arabie, P., & Boorman, S.A., Multidimensional scaling of measures of distance on partitions. *Journal of Mathematical Psychology,* 1973, 10, 148-203.

Arabie, P., & Soli, S.D., The interface between the type of regression and the methods of collecting proximities data. In Golledge, R., & Rayner, J.N., (Eds), *Multidimensional analysis of large data sets.* Columbus: Ohio State University Press, 1978.

Arnold, M.B., *Emotion and Personality* (2 volumes) New York: Columbia University Press, 1960.

Attneave, F., Dimensions of similarity. *American Journal of Psychology,* 1950, 63, 576-556.

Beals, R., & Krantz, D.H., Metrics and geodesics induced by order relations. *Mathematische Zeitschrift,* 1967, 101, 285-298.

Beals, R., Krantz, D.H., & Tversky, A., Foundations of multidimensional scaling. *Psychological Review,* 1968, 75, 127-42.

Bennett, J.F., & Hayes, W.L., Multidimensional scaling: determining the dimensionality of ranked preference data. *Psychometrika,* 1960, 25, 27-43.

Bentler, P.M., & Lee, S,Y., Statistical aspects of a three-mode factor analysis model. *Psychometrika,* 1978, 43, 343-352.

Bentler, P.M., & Weeks, D.G., Restricted multidimensional scaling models. *Journal of Mathematical Psychology,* 1978, 17, 138-151.

Bergmann, G., & Spence, K.W., The logic of psychophysical measurement. *Psychological Review,* 1944, 51, 1-24.

Block, J., Studies in the phenomenology of the emotions. *Journal of Abnormal and Social Psychology,* 1957, 54, 358-363.

Bloxam, B., *Individual differences in multidimensional scaling.* (ETS Research Bulletin), Princeton, N.J.: Educational Testing Service, 1968.

Bloxam, B., Constrained multidimensional scaling in N spaces. *Psychometrika,* 1978, 43, 397-408.

Borg, I., Some basic concepts of facet theory. In Lingoes, J.C., et.al. (Eds), *Geometric representations of relational data.* Ann Arbor: Mathesis Press, 1977 (a).

Borg, I., Geometric representation of individual differences. In Lingoes, J.C., et.al. (Eds), *Geometric representations of relational data.* Ann Arbor: Mathesis Press, 1977 (b).

Borg, I., & Lingoes, J.C., *A direct transformational approach to multidimensional analysis of three way data matrices: theory and applications.* (Michigan Mathematical Psychology Program Technical Report MMPP-76-1). Ann Arbor: University of Michigan, 1976.

Borg, I., & Lingoes, J.C., *Non-metric multidimensional scaling with order constraints on the distances.* (Michigan Mathematical Psychology Program Technical Report MMPP-78-3). Ann Arbor: University of Michigan, 1978, (a).

Borg, I., & Lingoes, J.C., What weights should weights have in individual differences scaling? *Quality and Quantity,* 1978, 12, 223-237 (b).

Borg, I., & Lingoes, J.C., Multidimensional scaling with side constraints on the distances. In Lingoes, J.C., et.al (Eds) *Geometric representations of relational data.* (2nd ed), Ann Arbor: Mathesis Press, 1979.

Borg, I., & Lingoes, J.C., A model and algorithm for multidimensional scaling with side constraints on the distances. *Psychometrika,* 1980, 45, 000-000.

Burton, M., Semantic dimensions of occupation names. In Shepard, R.N., et.al (Eds). *Multidimensional Scaling: theory and applications in the behavioural sciences.* Vol. II. New York: Seminar Press, 1972.

Bush, R.R., & Mosteller, F., *Stochastic models for learning.* New York: Wiley, 1955.

Campbell, N.R., *Physics: the elements.* Cambridge: Cambridge University Press, 1920.

Campbell, N.R., *What is science?* 1921. Reprinted New York: Dover Publications, 1952.

Campbell, N.R., *An account of the principles of measurement and calculation.* London: Longmans, Green, 1928.

Campbell, N.R., *Symposium: Measurement and its importance for philosophy.* Aristotelian Society, Suppl. Vol. 17, London: Harrison, 1938.

Carmone, F.J., Green, P.E., & Robinson, P.J., TRICON - An IBM 360/65 Fortran IV Program for the triangularization of conjoint data. *Journal of Marketing Research,* 1970, 5, 219-220.

Carroll, J.D., Individual differences and multidimensional scaling. In Shepard, R.N., et.al. (Eds), *Multidimensional scaling: theory and applications in behavioural sciences.* Vol. 1. New York: Seminar Press, 1972.

Carroll, J.D., Spatial, non-spatial and hybrid models for scaling. *Psychometrika,* 1976, 41, 439-463.

Carroll, J.D., & Chang, J.J., Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition *Psychometrika,* 1970, 35, 283-319 (a).

Carroll, J.D., & Chang, J.J., A reanalysis of some color data of Helm's by the INDSCAL procedure for individual differences scaling. *Proceedings, 78th annual convention, American Psychological Association,* 1970, 5, 137-138 (b).

Carroll, J.D., & Chang, J.J., *IDIOSCAL (Individual Differences in Orientation Scaling) A generalisation of INDSCAL allowing IDIOsyncratic reference system as well as an analytic approximation to INDSCAL.* Paper presented at meeting of the Psychometric Society, Princeton, N.J. March, 1972.

Carroll, J.D., & Chang, J.J., *Some methodological advances in INDSCAL.* Paper presented at the Spring meeting of the Psychometric Society, Stanford, California, April 1974.

Carroll, J.D., Green, P.E., & Carmone, F.J., *CANDELINC: A new method for multidimensional analysis with constrained solutions.* Paper presented at meeting of International Congress of Psychology, Paris, France, July, 1976.

Carroll, J.D., & Pruzansky, S., *MULTILINC: Multiway Candelinc.* Paper presented at the meeting of the American Psychological Association, San Francisco, August, 1977.

Carroll, J.D., & Wish, M., Models and methods for three-way multidimensional scaling. In Krantz, D.H., et.al. (eds), *Contemporary developments in mathematical psychology,* Vol. II, San Francisco: Freeman, 1974.

Catell, R.B., ^{r}p and other coefficients of pattern similarity. *Psychometrika,* 1949, 14, 279-298.

Cliff, N., Adjective checklist responses and individual differences in perceived meaning. *Educational and Psychological Measurement,* 1968, 28, 1063-1077 (a).

Cliff, N., The idealized individual interpretation of individual differences in multidimensional scaling. *Psychometrika,* 1968, 33, 225-232.(b).

Cliff, N., Liking judgements and multidimensional scaling. *Educational and Psychological Measurement,* 1972, 29, 73-86.

Cliff, N., Scaling. *Annual Review of Psychology,* 1973, 24, 473-506.

Cliff, N., & Young, F.W., Relation between unidimensional judgements and multidimensional scaling. *Organisational Behaviour and Human Performance* 1968, 3, 269-85.

Coombs, C.H., Psychological scaling without a unit of measurement. *Psychological Review,* 1950, 57, 145-58.

Coombs, C.H., *Theory of Data.* New York: Wiley, 1964.

Coombs, C.H., *Models and methods for the study of chemoreception-hedonics.* Paper presented to Third European Chemoreception Research Organisation Symposium on Chemoreception and Preference Behaviour, Netherlands, May, 1979.

Coombs, C.H., Raiffa, H., & Thrall, R.M., Some views on mathematical models and measurement theory. *Psychological Review,* 1954, 61, 132-44.

Cohen, M.R., & Nagel, E., *An introduction to logic and scientific method.* New York: Harcourt, Brace, 1934.

Cronbach, C.J., & Gleser, G.C., Assessing similarity between profiles. *Psychological Bulletin,* 1953, 50, 456-473.

Cronbach, C.J., Gleser, G.C., Nanda, H., & Nageswan, R., *The dependability of behavioural measurements: theory of generalizability for scores and profiles.* New York: Wiley, 1972.

Cunningham, J.P., & Shepard, R.N., Monotone mapping of similarities into a general metric space. *Journal of Mathematical Psychology,* 1974, 11, 335-363.

Davitz, J.R., *The language of emotion.* New York: Academic Press, 1969.

de Leeuw, J., *Non-metric discriminant analysis.* (R.N. 007-68) University of Leiden, 1968.

de Leeuw, J., & Pruzansky, S., A new computational method to fit the weighted Euclidean model. *Psychometrika,* 1978, 43, 479-490.

de Leeuw, J., Young, F.W., & Takane, Y., Additive structures in qualitative data. An alternating least-squares method with optimal scaling features. *Psychometrika,* 1976, 41, 471-503.

Eckart, C., & Young, G., Approximation of one matrix by another of lower rank. *Psychometrika,* 1936, 1, 211-218.

Eisler, H., On psychophysics in general and the general psychophysical differential equation in particular. *Scandinavian Journal of Psychology,* 1965, 6, 85-102.

Eisler, H., & Roskam, E.E., Multidimensional similarity: an experimental and theoretical comparison of vector, distance and set theoretical models. I. Models and internal consistency of data. *Acta Psychologica,* 1977, 41, 1-46, (a).

Eisler, H., & Roskam, E.E., Multidimensional similarity: an experimental and theoretical comparison of vector, distance and set theoretical models. II. Multidimensional analyses: the subjective space. *Acta Psychologica,* 1977, 41, 335-363. (b).

Ekman, G., Dimensions of color vision. *Journal of Psychology,* 1954, 38, 467-474.

Ekman, G., & Sjöberg, L., Scaling. *Annual Review of Psychology,* 1965, 16, 451-474.

Ellis, B.D., Some fundamental problems of direct measurement. *Australian Journal of Philosophy,* 1960, 38, 37-47.

Ellis, B.D., Some fundamental problems of direct measurement. *Australian Journal of Philosophy,* 1960, 38, 37-47.

Ellis, B.D., *Basic concepts of measurement.* Cambridge: Cambridge University Press, 1966.

Estes, W.K., The statistical approach to learning theory. In, Koch, S., (Ed) *Psychology: a study of a science,* Vol. 2, New York: McGraw-Hill, 1959.

Falmagne, J.C., Foundations of Fechnerian psychophysics. In Krantz, D.H., et.al. (Eds). *Contemporary developments in mathematical psychology.* Vol. II, San Francisco: Freeman, 1974.

Falmagne, J.C., Random conjoint measurement and loudness summation. *Psychological Review,* 1976, 83, 65-79.

Feger, H., Die erfassung individueller einstellungsstrukturen. *Zeitschrift fur Sozialpsychologie,* 1974, 5, 57-105.

Fenker, R.M., *Two contemporary problems in multidimensional scaling.* Technical Memorandum 8-72. Human Engineering Laboratory, Aberdeen Proving Ground, Md. 1972.

Feigenbaum, E.A., The simulation of verbal behaviour. *Proceedings of the Western Joint Computer Conference,* 1961, 19, 121-132.

Ferguson, A., et.al, Quantitative estimates of sensory events. Final report of the Committee appointed to consider and report on the possibility of Quantitative Estimates of Sensory Events. *Advancement of Science,* 1940, 1, 331-349.

Fillenbaum, S., & Rapoport, A., *Structures in the subjective lexicon.* New York: Academic Press, 1971.

Fisher, R.A., *Statistical methods for research workers.* (10th Ed), Edinburgh: Oliver and Boyd, 1946.

Flament, C., *Applications of graph theory to group structure.* Englewood Cliffs, N.J.: Prentice-Hall, 1963.

Foa, U.G., The contiguity principle in the structure of interpersonal relations. *Human Relations,* 1958, 11, 229-238.

Foa, U.G., The structure of interpersonal behaviour in the dyad. In Crisswill, J.H., et.al., (Eds) *Mathematical methods in small group processes.* Stanford, C.A.: Stanford University Press, 1962, 166-179.

Foa, U.G., New developments in facet design and analysis. *Psychological Review,* 1965, 72, 262-274.

Forgas, J.P., & Menyhart, J., The perception of political leaders: a multi-dimensional analysis. *Australian Journal of Psychology,* 1979, 31, 213-223.

Fraser, C.O., Cognitive strategies and multidimensional scaling. *British Journal of Psychology,* 1976, 67, 399-406.

Fraser, C.O., Measurement in psychology. *British Journal of Psychology,* 1980, 71, 23-34.

Goldmeier, E., Similarity in visually perceived forms. *Psychological Issues,* Monograph 29, 1972, 8, 1-136. (Originally published 1933).

Gower, J.C., Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika,* 1966, 53, 325-338.

Gower, J.C., Generalised Procrustes analysis. *Psychometrika*, 1975, 40, 33-51.

Gratch, H., (ed), *Twenty-five years of social research in Israel*. Jerusalem: Jerusalem Academic Press, 1973.

Green, P.E., & Carmone, F.J., *Multidimensional scaling and related techniques in marketing analysis*. Boston: Allyn & Bacon, 1970.

Green, P.E., & Carmone, F.J., The effect of task on intraindividual differences in similarity judgements. *Multivariate Behavioural Research,* 1971, 4, 433-450.

Green, P.E., & Rao, V.R., *Applied Multidimensional Scaling*. Hindsdale, Ill: Drysdale Press, 1972.

Greeno, J.G., Representation of learning as discrete transition in a finite state space. In Krantz, D.H., et.al, (Eds), *Contemporary developments in mathematical psychology*. *Vol. I.* San Francisco: Freeman, 1974.

Gregson, R.A.M., Odour similarities and their multidimensional metric representation. *Multivariate Behavioural Research,* 1972, 7, 165-174.

Gregson, R.A.M., *Psychometrics of Similarity.* New York: Academic Press, 1975.

Gregson, R.A.M., A comparative evaluation of seven similarity models. *British Journal of Mathematical and Statistical Psychology,* 1976, 29, 139-156.

Gregson, R.A.M., & Mitchell, M., Odor quality similarity scaling and odor-word profile matching. *Chemical Senses and Flavour,* 1974, 1, 95-101.

Gulliksen, H., A least squares solution for paired comparisions with incomplete data. *Psychometrika,* 1956, 21, 125-134.

Guttman, L., A basis for scaling qualitative data. *American Sociological Review,* 1944, 9, 139-50.

Guttman, L., An outline of some new methodology for social research, *Public Opinion Quarterly,* 1954-5, 18, 395-404.

Guttman, L., Introduction to facet design and analysis. In *Proceedings of Fifteenth International Congress of Psychology, Brussels.* Amsterdam: North-Holland Publishing, 1957.

Guttman, L., A structural theory for intergroup beliefs and actions. *American Sociology Review,* 1959, 24, 318-328.

Guttman, L., Measurement as structural theory. *Psychometrika,* 1971, 36, 329-347.

Harlow, H., & Stagner, R., Psychology of feelings and emotions II. Theory of emotions. *Psychological Review,* 1933, 40, 184-94.

Helm, C.E., Multidimensional ratio scaling analysis of perceived color relations. *Journal of the Optical Society of American,* 1964, 54, 256-262.

Helm, C.E., & Tucker, L.R., Individual differences in the structure of color perception. *American Journal of Psychology,* 1962, 75, 437-444.

Helmholtz, H.V., Zahlen und messen erktennis - theoretische betrachtet. In Philosophische aufsatze. Eduard Zeller gewidmet, Leipzig, 1887. Reprinted in *Gesammelte Abhandlungen,* Vol. 3, 1895, 356-391.

Hempel, C.G., Fundamentals of concept formation in empirical science. *International Encyclopedia of Unified Science, Vol. II, No. 7,* Chicago: University of Chicago Press, 1952.

Hendel, D.D., Intransitivity on paired-comparisons instruments: the relationship of the total circular triad score to stimulus circular triads. *Applied Psychological Measurement,* 1977, 1, 403-411.

Henley, N.M., Psychological study of the semantics of animal terms. *Journal of Verbal Learning Behaviour,* 1969, 8, 176-84.

Herzberg, F., Mausner, B., & Snyderman, B., *The motivation to work.* (2nd ed). New York: Wiley, 1959.

Hölder, O., Die axiome der quantität die lehre von mass. *Berichte urber die verhandlungen der Königlich sachsischen gessellschaft der wissenschaften zu Leipzig, Mathematisch - Physische Class,* 1901, 53, 1-64.

Holman, E.W., Completely non-metric multidimensional scaling. *Journal of Mathematical Psychology,* 1978, 18, 39-51.

Horan, C.B., Multidimensional scaling: combining observations when individuals have different perceptual structures. *Psychometrika,* 1969, 34, 139-65.

Howard, J.H., & Silverman, E., A multidimensional scaling analysis of 16 complex sounds. *Perception & Psychophysics,* 1976, 19, 193-200.

Hubert, L.J., & Subkoviak, M.J., Confirmatory inference and geometric models. *Psychological Bulletin,* 1979, 86, 361-370.

Isaac, P.D., Dissimilarities as indices of individual perceptual structure. *Perception & Psychophysics,* 1970, 7, 229-233.

Johnson, R.M., Pairwise non-metric multidimensional scaling. *Psychometrika,* 1973, 38, 11-18.

Johnson, S.C., Hierachical clustering schemes. *Psychometrika,* 1967, 32, 241-254.

Jones, R.A., & Ashmore, R.D., The structure of intergroup perception: categories and dimensions in views of ethnic groups and adjectives used in stereotype research. *Journal of Personality and Social Psychology,* 1973, 25, 428-438.

Jöreskog, K.G., A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika,* 1969, 34, 183-202.

Jöreskog, K.G., A general model for analysis of covariance structures. *Biometrika,* 1970, 57, 239-251.

Jöreskog, K.G., Statistical analysis of sets of congeneric tests. *Psychometrika,* 1971, 36, 109-133.

Just, M.R., & Murray, C.A., Community participation and development projects in Thai villages. In Shye, S., (Ed), *Theory construction and data analysis in the behavioural sciences.* San Francisco: Jossey-Bass, 1978.

Kaiser, H.F., A second generation little jiffy. *Psychometrika,* 1970, 35, 401-415.

Kaiser, H.F., & Serlin, R.C., Contributions to the method of paired comparisons. *Applied Psychological Measurement,* 1978, 2, 421-430.

Kehoe, J., & Reynolds, T.J., Interactive multidimensional scaling of cognitive structure underlying person perception. *Applied Psychological Measurement,* 1977, 1, 155-169.

Khincin, A.I., *Mathematical foundations of information theory.* New York: Dover, 1957.

Koopman, R.F., & Cooper, M., *Some problems with Minkowski distance models in multidimensional scaling.* Presented at annual meeting of the Psychometric Society, Stanford Univerisity, March 28, 1974.

Krantz, D.H., Rational distances functions for multidimensional scaling. *Journal of Mathematical Psychology*, 1967, 4, 226-245.

Krantz, D.H., A theory of context effects based on cross-context matching. *Journal of Mathematical Psychology*, 1968, 5, 1-48 (a).

Krantz, D.H., A survey of measurement theory. In Dantzig, G.B., & Veinott, A.F., Jr., (Eds), *Mathematics of the decision sciences, Part 2. Lectures in applied mathematics. Vol. 12.* Providence: American Mathematics Society, 1968, (b).

Krantz, D.H., A theory of magnitude estimation and cross-modality matching. *Journal of Mathematical Psychology*, 1972, 9, 168-199, (a).

Krantz, D.H., Measurement structures and psychological laws. *Science*, 1972, 175, 1427-1435, (b).

Krantz, D. H., Measurement theory and qualitative laws in psychophysics. In Krantz, D.H., et.al. (Eds), *Contemporary developments in mathematical psychology. Vol. II.* San Francisco: Freeman, 1974.

Krantz, D.H., Luce, R.D., Suppes, P., & Tversky, A., *Foundationsof measurement.* Vol. I. New York: Academic Press, 1971.

Krantz, D.H., & Tversky, A., Conjoint measurement analysis of composition rules in psychology. *Psychological Review*, 1971, 78, 151-159.

Krantz, D.H., & Tversky, A., Similarity of rectangles: an analysis of subjective dimensions. *Journal of Mathematical Psychology*, 1975, 12, 4-34.

Kruskal, J.B., Multidimensional scaling by optimizing goodness of fit to a non-metric hypothesis. *Psychometrika,* 1964, 29, 1-27, (a).

Kruskal, J.B., Non-metric multidimensional scaling: a numerical method. *Psychometrika,* 1964, 29, 28-42, (b).

Kruskal, J.B., & Shepard, R.N., A non-metric variety of linear factor analysis. *Psychometrika,* 1974, 39, 123-157.

Kruskal, J.B., Young, F.W., & Seery, J.B., *How to use KYST, a very flexible program to do multidimensional scaling.* (Unpublished Manuscript). Murray Hill, N.J.: Bell Laboratories, 1973.

LaPorte, R.E., & Voss, J.F., Prose representation: a multidimensional scaling approach. *Multivariate Behavioural Research,* 1979, 14, 39-56.

Lehner, P., & Noma, E., *A new solution to the problem of finding all numerical solutions to ordered metric structures.* (Michigan Mathematical Psychology Program Technical Report MMPP-79-2) Ann Arbor: University of Michigan, 1979.

Lemke, E., & Wiersman, W., *Principles of psychological measurement.* Chicago: Rand McNally, 1976.

Levelt, W.J.M., van der Geer, J.P., & Plomp, R., Triadic comparisions of musical intervals. *British Journal of Mathematical and Statistical Psychology,* 1966, 19, 163-179.

Lindman, H., & Caeli, T., Constant curvature Riemannian scaling. *Journal of Mathematical Psychology,* 1978, 17, 89-109.

Lingoes, J.C., The multivariate analysis of qualitative data. *Multivariate Behavioural Research,* 1968, 3, 61-94.

Lingoes, J.C., Identifying directions in the space for interpretation. In Lingoes, J.C., et.al. (Eds), *Geometric representations of relational data*. Ann Arbor: Mathesis Press, 1977, (a).

Lingoes, J.C., Identifying regions in the space for interpretation. In Lingoes, J.C., et.al. (Eds), *Geometric representations of relational data*. Ann Arbor: Mathesis Press, 1977, (b).

Lingoes, J.C., *Progressively complex linear transformations for finding geometric similarities among data structures*. Invited paper: Conference on Classifying Cultural and Social Data, Charleston, S.C., March 1977, (c).

Lingoes, J.C., *Testing regional hypotheses in multidimensional scaling*. Invited address: Second International Symposium on Data Analysis and Informatics, Versailles, France, October 17-19, 1979.

Lingoes, J.C., & Borg, I., Procrustean individual differences scaling: PINDIS. *Journal of Marketing Research,* 1976, 13, 406-407, (a).

Lingoes, J.C., & Borg, I., *Optimal solutions for dimension and vector weights in PINDIS*. (Michigan Mathematical Psychology Program Technical Report MMPP-76-4), Ann Arbor: University of Michigan, 1976, (b).

Lingoes, J.C., & Borg, I., Procrustean individual differences scaling: PINDIS. *Sozialpsychologie,* 1977, 8, 210-217, (a).

Lingoes, J.C., & Borg, I., Identifying spatial manifolds for interpretation. In Lingoes, J.C., et.al. (Eds), *Geometric representations of relational data*. Ann Arbor: Mathesis Press, 1977, (b).

Lingoes, J.C., & Borg, I., CMDA-U. Confirmatory monotone distance analysis - Unconditional. *Journal of Marketing Research,* 1978, 15, 610-611, (a).

Lingoes, J.C., & Borg, I., A direct approach to individual differences scaling using increasingly complex transformations. *Psychometrika,* 1978, 43, 491-519 (b).

Lingoes, J.C., & Borg, I., An exact significance test for dimensionality and choosing between rival hypotheses in confirmatory multidimensional scaling. *Psychometrika,* 1980, 45, 000-000.

Lingoes, J.C., & Schönemann, P.H., Alternative measures of fit for the Schöneman-Carroll matrixfitting algorithm. *Psychometrika,* 1974, 39, 423-427.

Lissitz, R.W., & Robinson, E.J., An examination of some factors related to using different Minkowski models in nonmetric multidimensional scaling. *Multivariate Behavioural Research,* 1977, 12, 69-74.

Luce, R.D., *Individual choice behaviour: a theoretical analysis.* New York: Wiley, 1959.

Luce, R.D., What sort of measurement is psychophysical measurement. *American Psychologist,* 1972, 27, 96-106.

Luce, R.D., Bush, R.R., & Galanter, E., *Handbook of mathematical psychology. I.* New York: Wiley, 1963.

Luce, R.D., & Tukey, J., Simultaneous conjoint measurement: a new type of fundamental measurement. *Journal of Mathematical Psychology,* 1964, 1, 1-27.

MacCallum, R.C., Relations between factor analysis and multidimensional scaling. *Psychological Bulletin,* 1974, 81, 505-516.

MacCallum, R.C., Effects on INDSCAL of nonorthogonal perceptions of object space dimensions. *Psychometrika,* 1976, 41, 177-188, (a).

MacCallum, R.C., Transformation of a three-mode multidimensional scaling solution to INDSCAL form. *Psychometrika,* 1976, 41, 385-400, (b).

MacCallum, R.C., Effects of conditionality on INDSCAL and ALSCAL weights. *Psychometrika,* 1977, 42, 297-305.

McClelland, G.H., & Coombs, C.H., ORDMET: A general algorithm for clustering all numerical solutions to ordered metric structures. *Psychometrika,* 1975, 40, 269-290.

McDonald, R.P., A simple comprehensive model for the analysis of covariance structures. *The British Journal of Mathematical and Statistical Psychology,* 1978, 31, 59-72.

McDonald, R.P., & Leony, K.S., *COSA: A FORTRAN programme for confirmatory factor analysis.* (Unpublished manuscript), Toronto: Ontario Institute for Studies in Education, 1976.

Mach, E., Die ahlichkeit und die analogie als leitmotiv der forschung. *Annal Naturphilos,* 1902, 1, 5-14.

Mashour, M., & Hosman, J., On the new "psychophysical law": a validation study. *Perception & Psychophysics,* 1968, 3, 367-375.

Maslow, A.H., *Motivation and personality.* New York: Harper & Row, 1954.

Messick, S., An empirical evaluation of multidimensional successive intervals. *Psychometrika,* 1956, 21, 367-375.

Micko, H.C., & Fischer, J., The metric of multidimensional psychological spaces as a function of the differential attention to subjective attributes. *Journal of Mathematical Psychology,* 1970, 7, 118-143.

Miller, G.A., A psychological method to investigate verbal concepts. *Journal of Mathematical Psychology.* 1969, 6, 169-191.

Miller, G.&., & Johson-Laird, P.N., *Language and perception.* Cambridge, Mass: Belknap, 1976.

Miller, G.A., & Nicely, P.E., An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America,* 1955, 27, 338-352.

Newell, A., & Simon, H.A., *Human problem solving.* Englewood Cliffs, N.J.: Prentice-Hall, 1972.

Noma, E., *Intersession variability and stress.* (Michigan Mathematical Psychology Program Technical Report MMPP-78-4). Ann Arbor: University of Michigan, 1978.

Noma, E., & Johnson, J., *Constraining nonmetric multidimensional scaling configurations*. (Technical Report No. 60). Ann Arbor, University of Michigan, Human Performance Centre, 1977.

Noma, E., & Johnson, J., *Constrained nonmetric multidimensional scaling*. (Michigan Mathematical Psychology Program Technical Report MMPP-79-4) Ann Arbor: University of Michigan, 1979.

Nunnally, J.C., *Psychometric theory*. New York: McGraw-Hill, 1967.

Nunnally, J.C., *Introduction to psychological measurement*. New York: McGraw-Hill, 1970.

Nygren, T.E., *A procedure for testing the axioms of the additive difference model*. Presented to meeting of Mathematical Society, Providence RI, August, 1979.

Offenbach, S.I., *Children's perception of Munsell colours*. Paper presented to Biennial Meeting of Society for Research in Child Development. San Francisco, 1979.

O'Hare, D., Individual differences in perceived similarity and preference for visual art: a multidimensional scaling analysis. *Perception and Psychophysics*, 1976, 20, 445-452.

Osgood, C.E., Dimensionality of the semantic space for communication via facial expressions. *Scandinavian Journal of Psychology*, 1966, 7, 1-30.

Pearson, K., On the coefficient of racial likeness. Biometrika, 1928, 18, 105-117.

Pfanzagl, J., *Theory of measurement.* Physica-Velag, 1968.

Piesko, H., Multidimensional scalings in Riemannian space. *Journal of Mathematical Psychology,* 1975, 12, 449-477.

Pruzansky, S., *SINDSCAL: A computer program for individual differences in multidimensional scaling.* (Unpublished manuscript). Princeton, N.J.: Bell Laboratories, 1975.

Prytulak, L.S., Critique of S.S. Steven's theory of measurement scale classification. *Perceptual Motor Skills,* 1975, 41, 3-38.

Rao, C.R., The utilization of multiple measurements in problems of biological classification. *Journal of the Royal Statistical Society, Section B.* 1948, 10, 159-203.

Rao, V.R., & Katz, R., Alternative multidimensional scaling methods for large stimulus sets. *Journal of Marketing Research,* 1971, 8, 488-494.

Restle, F. A metric and an ordering on sets. *Psychometrika,* 1959, 24, 207-20.

Ronen, S., Kraut, A.I., Lingoes, J.C., & Aranya, N. A nonmetric scaling approach to taxonomies of employee work motivation. *Multivariate Behavioural Research,* 1979, 14, 387-401.

Rosenberg, S. & Kim, M.P. The method of sorting as a data gathering procedure in multivariate research. *Multivariate Behavioural Research,* 1975, 10, 489-502.

Rosenberg, S. & Sedlak, A. Structural representation of perceived personality trait relationships. In Shepard, R.N., et al (Eds.) *Multidimensional Scaling: theory and applications in the behavioural sciences.* Vol.II, New York: Seminar Press, 1972.

Roskam, E., *Metric analysis of ordinal data in psychology.* Voorschoten, Holland: University of Leiden Press, 1968.

Roskam, E., The method of triads for nonmetric multidimensional scaling. *Nederlands Tijdschrift voor Psychologie,* 1970, 25, 404-417.

Roskam, E. Multidimensional scaling by metric transformation of data. *Nederlands Tijdschrift voor Psychologie,* 1972, 27, 486-508.

Roskam, E. The nature of data: interpretation and representation. An introduction to the theory of data. In Lingoes, J.C., et al (Eds.) *Geometric representations of relational data.* Ann Arbor: Mathesis Press, 1977.

Rösler, F. Identifying interindividual judgement differences. INDSCAL or three-mode factor analysis? *Multivariate Behavioural Research,* 1979, 14, 145-167.

Ross, S. *Logical foundations of measurement. A study in the philosophy of science.* Copenhagen: Scandinavian University Books, 1964.

Ross, J. & Cliff, N. A generalisation of the interpoint distance model. *Psychometrika,* 1964, 29, 167-176.

Rounds, J.B., Miller, T.W. & Dawis, R.V., Comparability of multiple rank order and paired comparisons methods. *Applied Psychological Measurement,* 1978, 2, 413-420.

Royce, J.R. (Ed.) *Multivariate analysis and psychological theory.* London: Academic Press, 1973.

Rotter, J.B., Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs,* 1966, 80, (1, Whole No. 609).

Rumelhart, D.E. & Abrahamson, A.A. A theory of analogical reasoning. *Cognitive Psychology,* 1973, 5, 1-28.

Russell, B. *The principles of mathematics.* (2nd. Ed.) New York: Norton, 1937.

Sachs, H. *Die enstehung der raumvorstellung aus sinnesempfindungen.* Breslau: Schlettersche Buchh, 1897.

Schönemann, P.H., A note on Gulliksen's least squares solutions. *British Journal of Mathematical and Statistical Psychology,* 1970, 23, 69-71.

Schonemann, P.H., *An algebraic solution for a class of subjective metrics models.* Paper presented to spring meetings of Psychometric Society, St. Louis, April, 1971.

Schonemann, P.H. An algebraic solution for a class of subjective metrics models. *Psychometrika,* 1972, 37, 441-451.

Schönemann, P.H., Carter, F.S., & James, W.L. *Contributions to subjective metrics scaling. I. COSPA, a fast method for fitting and testing Horan's model, and an empirical comparison with INDSCAL and ALSCAL.* (Institute Paper No. 587). Purdue University, Krannert Graduate School, 1976.

Schönemann, P.H., James, W.L. & Carter, F.S. COSPA, Common Space Analysis - a program for fitting and testing Horan's subjective metrics model. *Journal of Marketing Research,* 1978, 15, 268-270.

Schönemann, P.H., James, W.L., & Carter, F.S. Statistical inference in multidimensional scaling: a method for fitting and testing Horan's model. In Lingoes, J.C. et al (Eds.) *Geometric representations of relational data.* (2nd Ed.) Ann Arbor: Mathesis Press, 1979.

Scott, D. & Suppes, P. Foundation aspects of theories of measurement. *Journal of Symbolic Logic,* 1958, 23, 113-128.

Shepard, R.N. Stimulus and response generalization: a stochastic model relating generalization to distance in psychological space. *Psychometrika,* 1957, 22,325-345.

Shepard, R.N. Stimulus and response generalization: tests of a model of relating generalization to distance in psychological space. *Journal of Experimental Psychology.* 1958, 55, 509-23(a).

Shepard, R.N. Stimulus and response generalization: deduction of the generalization gradient from a trace model. *Psychological Review,* 1958, 65, 242-256(b).

Shepard, R.N. The analysis of proximities: multidimensional scaling with an unknown distance function I. *Psychometrika,* 1962, 27, 127-140(a).

Shepard, R.N. The analysis of proximities: multidimensional scaling with an unknown distance function II. *Psychometrika,* 1962, 27, 219-246(b).

Shepard, R.N. Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology,* 1964, 1, 54-87.

Shepard, R.N. Metric structures in ordinal data. *Journal of Mathematical Psychology,* 1966, 3, 287-315.

Shepard, R.N. A taxonomy of some principal types of data and of multidimensional methods for their analysis. In Shepard, R.N., Romney, A.K. & Nerlove, S.B. (Eds.) *Multidimensional Scaling: theory and applications in the behavioural sciences.* Vol.I. New York: Seminar Press, 1972.

Shepard, R.N., Representation of structure in similarity data: problems and prospects. *Psychometrika,* 1974, 39, 373-421.

Shepard, R.N. The circumplex and related topological manifolds in the study of perception. In Shye, S. (Ed.) *Theory construction and data analysis in the behavioural sciences.* San Francisco: Jossey-Bass, 1978.

Shepard, R.N. & Arabie, P. Additive clustering: representation of similarities as combinations of discrete overlapping properties. *Psychological Review,* 1979, 86, 87-123.

Shepard, R.N. & Carroll, J.D. Parametric representation of nonlinear data structure. In Krishnaiah, P.R. (Ed.), *International symposium on multivariate analysis, Dayton, Ohio, 1965.* New York: Academic Press, 1966.

Shepard, R.N. & Metzler, J. Mental rotation of three-dimensional objects. *Science,* 1971, 171, 701-703.

Sherman, C.R. Nonmetric multidimensional scaling: a Monte Carlo study of the basic parameters. *Psychometrika,* 1972, 37, 323-355.

Shikiar, R. & Coates, C. A multidimensional scaling study of person perception in children. *Multivariate Behavioural Research* 1978, 13, 363-370.

Shye, S. On the search for laws in the behavioural sciences. In Shye, S. (Ed.) *Theory construction and data analysis in the behavioural sciences.* San Francisco: Jossey-Bass, 1978.(a).

Shye, S. Facet analysis and regional hypotheses. In Shye, S. (Ed.) *Theory construction and data analysis in the behavioural sciences.* San Francisco: Jossey-Bass, 1978.(b).

Shye, S. Achievement motive: a faceted definition and structural analysis. *Multivariate Behavioural Research,* 1978, 13, 327-346.(c).

Silver, C.A., Landis, D., & Messick, S. Multidimensional analysis of visual form: an analysis of individual differences. *American Journal of Psychology,* 1966, 79, 67-72.

Simon, H.A. & Newell, A. Thinking processes. In Krantz, D.H. et al (Eds.) *Contemporary developments in mathematical psychology*. Vol.II. San Francisco: Freeman, 1974.

Spence, I. & Domoney, D.W. Single subject incomplete designs for nonmetric multidimensional scaling. *Psychometrika*, 1974, 39, 469-490.

Spence, I. & Young, F.W. Monte Carlo studies in nonmetric scaling. *Psychometrika*, 1978, 43, 115-117.

Steffre, V. Some applications of multidimensional scaling to social science problems. In Shepard, R.N. et al (Eds.) *Multidimensional scaling: theory and applications in the behavioural sciences*. Vol.II New York: Seminar Press, 1972.

Stevens, S.S. On the problem of scales for the measurement of psychological magnitudes. *Journal for the Unification of Science*, 1939, 9, 94-99.

Stevens, S.S. On the theory of scales of measurement. *Science*, 1946, 103, 677-680.

Stevens, S.S. Mathematics, measurement and psychophysics. In Stevens, S.S. (Ed.) *Handbook of Experimental psychology*. New York: Wiley, 1951.

Stevens, S.S. The measurement of loudness. *Journal of the Acoustical Society of America*, 1955, 27, 815-820.

Stevens, S.S. The direct estimation of sensory magnitudes - loudness. *American Journal of Psychology*, 1956, 69, 1-25.

Stevens, S.S. On the psychophysical law. *Psychological Review*, 1957, 64, 153-183.

Stevens, S.S. The quantification of sensation. *Daedalus*, 1959, 88, 606-621(a).

Stevens, S.S. Cross-modality validation of subjective scales for loudness, vibration and electric shock. *Journal of Experimental Psychology*, 1959, 57, 201-209(b).

Stevens, S.S. On the new psychophysics, *Scandinavian Journal of Psychology*, 1960, 1, 27-35.

Stevens, S.S. Duration, luminance and the brightness exponent. *Perception & Psychophysics*, 1966, 1, 96-100 (a).

Stevens, S.S. Matching functions between loudness and ten other continua. *Perception & Psychophysics*, 1966, 1, 5-8(b).

Stevens, S.S. On the operation known as judgement. *American Scientist*, 1966, 54, 385-401(c).

Stevens, S.S. Measurement, statistics and the schemaparic view. *Science*, 1968, 161, 849-856.

Stevens, S.S. Neural events and the psychophysical law. *Science*, 1970, 170, 1043-1050.

Stevens, S.S. Issues in psychophysical measurement. *Psychological Review*. 1971, 5, 426-450.

Stevens, S.S.& Galanter, E.H. Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology,* 1957, 54, 377-411.

Suppes, P. & Zinnes, J.L. Basic measurement theory. In Luce, R.D., Bush, R.R. & Galanter, E.G. (Eds.) *Handbook of mathematical psychology.* Vol.I New York: Wiley, 1963.

Takane, Y., Young, F.W. & deLeuuw, J. Nonmetric individual differences multidimensional scaling: an alternating least squares method with optimal scaling features. *Psychometrika,* 1977, 42, 7-67.

Torgeson, W.S. Multidimensional Scaling. I. Theory and method. *Psychometrika,* 1952, 17, 401-419.

Torgerson, W.S., Theory and methods of scaling. New York: Wiley, 1958.

Torgerson, W.S. Multidimensional Scaling of similarity. *Psychometrika,* 1965, 30, 379-393.

Treisman, M. Noise and Weber's Law: the discrimination of brightness and other dimensions. *Psychological Review,* 1964, 71, 314-30.

Tucker, L.R. Some mathematical notes on three-mode factor analysis. *Psychometrika,* 1966, 31, 279-311.

Tucker, L.R. The extension of factor analysis to three-dimensional matrices. In Fredriksen, N. & Gulliksen, H. (Eds.) *Contributions to mathematical psychology.* New York: Holt, Rinehart & Winston, 1964.

Tucker, L.R., Relations between multidimensional scaling and three-mode factor analysis. *Psychometrika*, 1972, 37, 3-27.

Tucker, L.R. & Messick, S. An individual differences model for multidimensional scaling. *Psychometrika*, 1963, 28, 333-367.

Tukey, J.W. The future of data analysis. *Annals of Mathematical Statistics*, 1962, 33, 1-67.

Tversky, A. Features of similarity. *Psychological Review*, 1977, 84, 327-352.

Tversky, A. & Krantz, D.H. Similarity of schematic faces - a test of interdimensional additivity. *Perception & Psychophysics*, 1969, 5, 124-128.

Tversky, A & Krantz, D.H. The dimensional representation and the metric structure of similarity data. *Journal of Mathematical Psychology*, 1970, 7, 572-596.

Tzeng, O.C.S. & Landis, D. Three-mode multidimensional scaling with points of view solutions. *Multivariate Behavioural Research*, 1978, 13, 181-213.

Van der Kamp, L.J.T. & Pols, L.C.W., Perceptual analysis from confusions between vowells. *Acta Psychologica*, 1971, 35, 64-77.

Wagenaar, W.A. Application of Luce's choice axiom to form discrimination. *Nederlands Tijdschrift voor Psychologie*, 1968, 23, 96-108.

Waganaar, W.A. Stevens vs. Fechner: a plea for dismissal of the case. *Acta Psychologica*, 1975, 39, 225-235.

Ward, L.M. Multidimensional scaling of the molar physical environment. *Multivariate Behavioural Research*, 1977, 12, 23-42.

Weeks, D.G. & Bentler, P.M. A comparison of linear and monotone multidimensional scaling models. *Psychological Bulletin*, 1979, 86, 349-354.

Wender, K. *Interpretation of Minkowski metrics in ordinal multidimensional scaling*. Paper presented to Second Annual Meeting of Gruppe Mathematische Psychologie, Marburg/Lahn, April 1968.

Wiener-Ehrlich, W.K. Dimensional and Metric structures in multidimensional stimuli. *Perception & Psychophysics*, 1978, 24, 399-414.

Wilson, K.V. Multidimensional analysis of confusions of English consonants. *American Journal of Psychology*, 1963, 76, 89-95.

Wish, M. Compairsons among multidimensional structures of nations based on difference measures of subjective similarity. *General Systems*, 1970, 15, 55-65.

Wish, M. Individual differences in perception and preferences among nations. In King, C.W. & Tigert, D. (Eds.) *Attitude research reaches new heights*. Chicago: American Market Association, 1971.

Wish, M. Comparisons among multidimensional structures of interpersonal relations. *Multivariate Behavioural Research*, 1976, 11, 297-324.

Wish, M. & Carroll, J.D. Psychophysical judgement and measurement. In, Carterette, E.C. & Friedman, M.P. (Eds.) *Handbook of Perception,* Vol.II, New York: Academic Press, 1974.

Wish, M. Deutsch, M. & Biener, L. Differences in conceptual structures of nations: an exploratory study. *Journal of Personality and Social Psychology,* 1970, 16, 361-373.

Wish, M. Deutsch, M. & Biener, L. Differences in perceived similarity of nations. In Shepard, R.N., Romney, A.K. & Nerlove, S. (Eds.) *Multidimensional scaling: theory and applications in the behavioural sciences*. Vol.II. New York: Seminar Press, 1972.

Wish, M., Kaplan, S.J. & Deutsch, M. Dimensions of interpersonal relations: preliminary results. *Proceedings, 81st Annual convention, American Psychological Association,* 1973, 8, 179-180.

Wundt, W. Grundiss der psychologie, Leipzig, 1896 (No.109), Translated by C.H. Judd as *Outlines of psychology,* London and New York, 1897 (No.96).

Yoshida, M. Dimensions of tactual impressions. Parts A and B. *Japanese Psychological Research,* 1968, 10, 123-127, 157-173.

Young, F.W. Nonmetric multidimensional scaling: recovery of metric information. *Psychometrika,* 1970, 35, 455-473.

Young, F.W. *Conjoint scaling*. (Technical Report). Chapel Hill: L.L. Thurstone Psychometric Laboratory, 1973.

Young. F.W. Methods for describing ordinal data with cardinal models. *Journal of Mathematical Psychology*, 1975, 12, 416-436.

Young, F.W. *Principal directions scaling: a new individual differences model.* Presented to meeting of Psychometric Society, Monterey, C.A., June, 1979 (a).

Young, F.W. *Principal directions scaling: simultaneous analysis of similarity and multivariate data.* Presented to annual meeting of Society of Multivariate Experimental Psychology, Los Angeles, Oct., 1979(b).

Young, F.W. & Cliff, N. Interactive scaling with individual subjects. *Psychometrika*, 1972, 37, 385-415.

Young, F.W. deLeuuw, J., & Takane, Y. Multiple (and canonical) regression with a mix of qualitative and quantitative variables: an alternating least squares method with optimal scaling features. *Psychometrika*, 1976, 41, 505-529.

Young, F.W. & Null, C.H., Multidimensional scaling of nominal data: the recovery of metric informatin with ALSCAL. *Psychometrika*, 1978, 43, 367-380.

Young, F.W., Null, C.H. & Sarle, W. Interactive similarity ordering. *Behaviour Research Methods & Instrumentation*, 1978, 10, 273-280.

Young, F.W., & Torgerson,W.S. TORSCA, a FORTRAN IV program for nonmetric multidimensional scaling. *Behavioural Science*, 1968, 13, 343-344.

Young, G., & Householder, A.S., Discussion of a set of points in terms of their mutual distances. *Psychometrika,* 1938, 3, 19-22.

Young, P.T., *Motivation and emotion.* New York: Wiley, 1961.

Zinnes, J.L. Scaling. *Annual Review of Psychology,* 1969, 20, 447-478.

REFERENCE NOTES

1. Gregson, R.A.M., Similarities within a set of closed group transformations. Unpublished paper, University of Canterbury, 1972.

2. Guttman, L., Faretted SSA. Unpublished paper, Hebrew University of Jerusalem, 1976.

3. Skarabis, H., Multidimensional scaling with prescribed structures - I. Unpublished paper, Freie University, Berlin, Germany, (circa 1976).

GIPPSLAND INSTITUTE OF ADVANCED EDUCATION

INSTRUCTIONS

The following list contains pairs of words referring to emotions, that is, different types of feelings. For each pair you are to assign a number between 0 and 10, expressing how similar you think the emotions referred to are. '10' means they are the same, '0' means they are not at all similar. Any number between 0 and 10 is possible, the higher the number the more similar the two emotions appear to be.

Do not spend too much time on each pair, and do not go back over any pairs you have previously considered. Just write down opposite each pair the number that you feel expresses the similarity of the two emotions.

The following list gives all the emotion names that will be used in this questionnaire. Please scan this list and ask if there are any words you are unfamiliar with.

Contempt	Pride
Disgust	Worry
Humiliation	Anticipation
Anger	Guilt
Envy	Boredom
Love	Sadness
Fear	Sympathy
Joy	Grief

REMEMBER - THE MORE SIMILAR THEY APPEAR TO BE THE HIGHER THE NUMBER YOU SHOULD GIVE.

Thank you for your co-operation.

Anger
Worry ______

Disgust
Grief ______

Envy
Sympathy ______

Pride
Sadness ______

Envy
Disgust ______

Worry
Anticipation ______

Guilt
Grief ______

Sympathy
Grief ______

Envy
Sadness ______

Contempt
Envy ______

Humiliation
Joy ______

Humiliation
Pride ______

Anger
Love ______

Worry
Grief ______

Guilt
Fear ______

Disgust
Worry ______

Pride
Love ______

Guilt
Boredom ______

Joy
Love ______

Joy
Guilt ______

Guilt
Contempt ______

Envy
Fear ______

Humiliation
Disgust ______

Anger
Anticipation ______

Disgust
Boredom ______

Sadness
Anger ______

Fear
Worry ______

Anticipation
Fear ______

Love
Boredom ______

Envy
Joy ______

Pride
Sympathy ______

Sympathy
Disgust ______

Love
Guilt ______

Joy
Grief ______

Boredom
Fear ______

Contempt
Pride ______

Anticipation
Contempt ______

Love
Grief ______

Guilt
Humiliation ______

Sadness
Sympathy ______

Anger
Boredom ______

Disgust
Guilt ______

Love
Envy ______

Disgust
Contempt ______

Anticipation
Pride ______

Humiliation
Sympathy ______

Humiliation
Sadness ______

Worry
Guilt ______

Boredom
Envy ______

Worry
Humiliation ______

Pride
Fear ______

Sympathy
Guilt ______

Boredom
Joy ______

Anger
Guilt ______

Humiliation
Envy ______

Boredom
Humiliation ______

Sympathy
Worry ______

Disgust
Love ______

Contempt
Joy ______

Fear
Joy ______

Joy Anger	________	Anticipation Humiliation	________
Contempt Humiliation	________	Sadness Worry	________
Grief Anticipation	________	Grief Boredom	________
Sadness Disgust	________	Disgust Anger	________
Boredom Pride	________	Love Contempt	________
Anger Grief	________	Guilt Envy	________
Worry Pride	________	Love Humiliation	________
Anticipation Envy	________	Pride Joy	________
Contempt Sympathy	________	Grief Contempt	________
Envy Anger	________	Fear Sympathy	________
Sadness Guilt	________	Humiliation Fear	________
Sympathy Anger	________	Worry Contempt	________
Guilt Anticipation	________	Love Anticipation	________
Anticipation Boredom	________	Grief Humiliation	________
Anger Humiliation	________	Sympathy Anticipation	________

Joy
Disgust ________

Joy
Worry ________

Sadness
Joy ________

Anger
Contempt ________

Grief
Envy ________

Fear
Love ________

Joy
Sympathy ________

Grief
Sadness ________

Fear
Sadness ________

Boredom
Contempt ________

Sympathy
Love ________

Anticipation
Sadness ________

Contempt
Fear ________

Worry
Boredom ________

Fear
Anger ________

Fear
Disgust ________

Love
Worry ________

Pride
Disgust ________

Sadness
Love ________

Grief
Fear ________

Envy
Pride ________

Guilt
Pride ________

Boredom
Sympathy ________

Disgust
Anticipation ________

Worry
Envy ________

Pride
Anger ________

Grief
Pride ________

Boredom
Sadness ________

Anticipation
Joy ________

Contempt
Sadness ________